12-13-2008

Quantifying cognitive workload and defining training time requirements using thermography

Jihun Kang

QUANTIFYING COGNITIVE WORKLOAD AND DEFINING TRAINING TIME REQUIREMENTS USING THERMOGRAPHY

By

Jihun Kang

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in the Department of Industrial and Systems Engineering

Mississippi State, Mississippi

December 2008

QUANTIFYING COGNITIVE WORKLOAD AND DEFINING TRAINING TIME REQUIREMENTS USING THERMOGRAPHY

By

Jihun Kang

Approved:

____
Kari Babski-Reeves
Assistant Professor of
Industrial and Systems Engineering
(Director of Dissertation)

____
Burak Eksioglu
Assistant Professor of
Industrial and Systems Engineering
(Committee Member)

____
Lesley Strawderman
Assistant Professor of
Industrial and Systems Engineering
(Committee Member)

____
Gary M. McFadyen
Assistant Research Professor of
Center for Advanced Vehicular Systems
(Committee Member)

____
John M. Usher
Professor and Graduate Coordinator
of Industrial and Systems Engineering

____
Sarah A. Rajala
Dean of the College of Engineering

Name: Jihun Kang

Date of Degree: December 12, 2008

Institution: Mississippi State University

Major Field: Industrial and Systems Engineering

Major Professor: Dr. Kari Babski-Reeves

Title of Study: QUANTIFYING COGNITIVE WORKLOAD AND DEFINING TRAINING TIME REQUIREMENTS USING THERMOGRAPHY.

Pages in Study: 273

Candidate for Degree of Doctor of Philosophy

Effective mental workload measurement is critical because mental workload significantly affects human performance. A non-invasive and objective workload measurement tool is needed to overcome limitations of current mental workload measures. Further, training/learning increases mental workload during skill or knowledge acquisition, followed by a decreased mental workload, though sufficient training times are unknown.

The objectives of this study were to: (1) investigate the efficacy of using thermography as a non-contact physiological measure to quantify mental workload, (2) quantify and describe the relationship between mental workload and learning/training,

and, (3) introduce a method to determine a sufficient training time and an optimal human performance level for a novel task by using thermography.

Three studies were conducted to address these objectives. The first study investigated the efficacy of using thermography to quantity the relationship between mental workload and facial temperature changes while learning an alpha-numeric task. Thermography measured and quantified the mental workload level successfully. Strong and significant correlations were found among thermography, performance, and subjective workload measures (MCH and SWAT ratings).

The second study investigated the utility of using a psychophysical approach to determine workload levels that maximize performance on a cognitive task. The second study consisted of an adjustment session (participants adjusted their own workload levels) and work session (participants worked at the chosen workload level). Participants were found to fall into two performance groups (low and high performers by accuracy rate) and results were significantly different. Thermography demonstrated whether both group found their optimal workload level.

The last study investigated efficacy of using thermography to quantify mental workload level in a complex training/learning environment. Experienced drivers' performance data was used as criteria to indicate whether novice drivers mastered the driving skills. Strong and significant correlations were found among thermography, subjective workload measures, and performance measures in novice drivers.

This study verified that thermography is a reliable and valid way to measure workload as a non-invasive and objective method. Also, thermography provided more practical results than subjective workload measures for simple and complex cognitive tasks. Thermography showed the capability to identify a sufficient training time for simple or complex cognitive tasks.

DEDICATION

I would like to dedicate this dissertation to my God and my family: Mother, Father, Ji-in, Christie, and Edward.

ACKNOWLEDGEMENTS

I would like to express my appreciation to everyone who has supported and helped me during my long journey.

First of all, I would like to take this opportunity to extend my deepest gratitude to my advisor, Dr. Kari Babski-Reeves. She always supported and educated me as a great teacher and an ideal colleague. Without her encouragement and guidance, I might not have been able to finish my dissertation. I am honored to be her first Ph.D. student from MSU (Thank you Kari!!).

I would like to thank to Mr. John McGinley and Dr. Burak Eksioglu for their assistance and encouragement me before Dr. Babski-Reeves became my advisor. I sincerely thank my committee members, Dr. Lesley Strawderman and Dr. Gary McFadyen, for kindly taking over the task of being my committee members.

I would like to thank my friends: Chris Blackledge, Daniel Carruth, David Close, Shaheen Ahmed, Robin Littlejohn, Sourav Patnaik, Mark Thomas, Teena Garrison, Jacie Williams, Justin Smith, and Allison Robinson. They never failed to help me when I needed their hands and I enjoyed working with them. I also want to give thanks to the Brazzeal family for providing such a personnel relationship while living in Starkville.

Personally, I would like to thank to my wife, Ji-in Lee, the best gift from God for my life, for her deep understanding, patience, and incredible support during my entire period of study. I want to thank my parents and my children, Christie Kang and Edward Kang, for their unconditional love and belief.

Finally, I would like to thank my almighty God, always works for me.

TABLE OF CONTENTS

Page

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF TABLES ix

LIST OF FIGURES xiv

CHAPTER

1 INTRODUCTION 1

1.1 PROBLEM STATEMENT 5
1.2 OBJECTIVES 6
1.3 RESEARCH OUTLINE 7
1.4 REFERENCES 9

2 EFFICACY OF USING THERMOGRAPHY TO QUANTIFYING MENTAL WORKLOAD WHILE PERFORMING A NOVEL COGNITIVE TASK 13

2.1 ABSTRACT 13
2.2 INTRODUCTION 14
2.3 BACKGROUND 16
2.3.1 MENTAL WORKLOAD MEASUREMENT 16
2.3.2 THERMOGRAHY AS A MENTAL WORKLOAD MEASURE 23
2.4 OBJECTIVES 24
2.5 HYPOTHESES 25
2.6 METHODOLOGY 26
2.6.1 EXPERIMENTAL DESIGN 26
2.6.2 PARTICIPANTS 26
2.6.3 TASK DESCRIPTION 27
2.6.4 INDEPENDENT VARIABLES 27
2.6.5 DEPENDENT VARIABLES 28

2.6.6 PROCEDURE 31
2.6.7 DATA ANALYSIS 32
2.7 RESULTS 32
2.7.1 THERMOGRAPHY 34
2.7.2 SUBJECTIVE WORKLOAD MEASURES 37
2.7.3 PERFORMANCE MEASURES 39
2.7.4 RELATIONSHIP AMONG MEASURES 41
2.7.5 REGRESSION MODEL 44
2.8 DISCUSSION 46
2.8.1 TEST HYPOTHESES 47
2.8.2 SUBJECTIVE WORKLOAD ASSESSMENT COMPARISONS... 52
2.8.3 SUFFICIENT TRAINING TIME 53
2.8.4 LIMITATION OF SUBJECTIVE WORKLOAD AND PERFORMANCE MEASURES 55
2.9 FUTURE WORK AND LIMITATIONS 56
2.10 CONCLUSION 58
2.11 REFERENCES 59

3 QUANTIFICATION OF OPTIMAL MENTAL WORKLOAD AND HUMAN PERFORMANCE LEVELS BASED ON THERMAL IMAGING OF THE FACE 65

3.1 ABSTRACT 65
3.2 INTRODUCTION 66
3.3 BACKGROUND 67
3.3.1 PERFORMANCE AND MENTAL WORKLOAD 67
3.3.2 MENTAL WORKLOAD AND TIME CONSTRAINTS 72
3.3.3 MENTAL WORKLOAD AND THERMOGRAPHY 755
3.3.4 PSYCHOPHYSICAL ESTIMATION OF WORKLOADS/TASK DEMANDS 77
3.4 OBJECTIVE 79
3.5 HYPOTHESES 79
3.6 METHODOLOGY 80
3.6.1 EXPERIMENTAL DESIGN 80
3.6.2 PARTICIPANTS 80
3.6.3 TASK DESCRIPTION 81
3.6.4 INDEPENDENT VARIABLES 82
3.6.5 DEPENDENT VARIABLES 83
3.6.6 PROCEDURE 85
3.6.7 DATA ANALYSIS 85
3.7 RESULTS 86
3.7.1 OVERALL EXPERIMENT SESSION 87

3.7.1.1 ADJUSTMENT SESSION ANALYSES IN OVERALL SESSION 91
3.7.1.2 INITIAL PACE EFFECTS 94
3.7.1.3 WORK SESSION ANALYSES IN OVERALL SESSION 95
3.7.2 HIGH PERFORMING PARTICIPANTS AND LOW PERFORMING PARTICIPANTS........ 97
3.7.2.1 HIGH PERFORMER GROUP........ 98
3.7.2.1.1 HIGH PERFORMER ADJUSTMENT SESSION 99
3.7.2.1.2 HIGH PERFORMER WORK SESSION 102
3.7.2.2 LOW PERFORMER GROUP 103
3.7.2.2.1 LOW PERFORMER ADJUSTMENT SESSION 104
3.7.2.2.2 LOW PERFORMER WORK SESSION 106
3.8 DISCUSSION 108
3.8.1 TEST HYPOTHESES 109
3.8.2 WORKLOAD LEVEL OF HIGH PERFORMERS AND LOW PERFORMERS 117
3.8.3 USE OF PSYCHOPHYSICS TO DETERMINE OPTIMAL WORKLOAD LEVELS 120
3.9 FUTURE WORK........ 121
3.10 CONCLUSION........ 122
3.11 REFERENCES 123

4 ASSESSING TRAINING TIME AND MENTAL WORKLOAD DURING COMPLEX TASK PERFORMANCE USING THERMOGRAPHY........ 129

4.1 ABSTRACT........ 129
4.2 INTRODUCTION 130
4.3 BACKGROUD 132
4.3.1 TRAINING AND MENTAL WORKLOAD........ 132
4.3.2 MENTAL WORKLOAD MEASURE AND THERMOGRAPHY 134
4.3.3 SIMULATION AND DRIVER EDUCATION........ 135
4.3.4 NOVICE DRIVERS AND DRIVING SIMULATION........ 137
4.4 OBJECTIVES 138
4.5 HYPOTHESES........ 139
4.6 METHODOLOGY 140
4.6.1 EXPERIMENTAL DESIGN 140
4.6.2 PARTICIPANTS 141
4.6.3 TASK DESCRIPTION 141
4.6.4 INDEPENDENT VARIABLES 143
4.6.5 DEPENDENT VARIABLES........ 143
4.6.6 PROCEDURE........ 146
4.6.7 DATA ANALYSIS........ 146

4.7 RESULTS ... 149
4.7.1 EXPERIENCED DRIVERS ... 156
4.7.1.1 INDEPENDENT EVENTS IN EACH SCENARIO FOR EXPERIENCED DRIVERS ... 164
4.7.2 NOVICE DRIVER GROUP ... 179
4.7.2.1 INDEPENDENT EVENTS IN EACH SCENARIO ... 186
4.7.3 SKILL MASTERY AND LEARNING ... 202
4.7.4 THE TREND OF MENTAL WORKLOAD MEASURES ... 220
4.8 DISCUSSION ... 222
4.9 FUTURE WORK AND LIMITATIONS ... 238
4.10 CONCLUSIONS ... 240
4.11 REFERENCES ... 241

5 CONCLUSION ... 249

APPENDIX

A. MODIFIED COOPER-HARPER SCALE ... 251

B. SUBJECTIVE WORKLOAD ASSESSMENT TECHNIQUE (SWAT) ... 253

C. DEMOGRAPHIC AND HEALTH SCREENING FORM ... 255

D. PRE-EXPOSURE SIMULATOR SICKNESS FORM ... 258

E. POST-EXPOSURE SIMULATOR SICKNESS FORM ... 260

F. HIGH PERFORMERS’ THERMAL READINGS ... 262

G. LOW PERFORMERS’ THERMAL READINGS ... 270

LIST OF TABLES

2.1 Workload measurement methods 17

2.2 Mental workload measurement issues 19

2.3 Descriptive statistics for the dependent variables(values are in mean (standard deviation)) 33

2.4 Repeated measures ANOVA results (p-values) 34

2.5 Tukey's post hoc comparisons for ΔNT 36

2.6 Tukey's post hoc comparisons for subjective workload assessments 38

2.7 Tukey's post hoc comparisons for subjective workload assessments 40

2.8 Correlation coefficients for all significant correlations 42

2.9 Regression model performance 46

3.1 Descriptive statistics for the dependent variables (values are in mean (standard deviation)) 87

3.2 Mixed factors ANOVA results for the overall experimental session 88

3.3 Tukey's post hoc comparisons for each task block for the overall experimental session 90

3.4 Correlation coefficients for all significant correlations across all blocks 91

3.5 Mixed factors ANOVA results for the adjustment session 92

3.6 Tukey's post hoc comparisons for each task block in total adjustment session ... 93

3.7 Correlation coefficients for the adjustment session for overall session 93

3.8 Descriptive statistics and ANOVA results for initial pace effects 94

3.9 Mixed factors AVNOVA results for the work session 95

3.10 Tukey's post hoc comparisons for the work session 96

3.11 Correlation coefficients for work session 96

3.12 ANOVA results between high and low performers for the entire testing session 98

3.13 Descriptive statistics for high performers (values are in mean (standard deviation)) 99

3.14 Mixed factors ANOVA results for the adjustment session for high performers 100

3.15 Tukey's post hoc comparisons for the adjustment session for high performers. 101

3.16 Correlation coefficients for high performers' adjustment session 101

3.17 Mixed factors ANOVA results for high performers' work session 102

3.18 Correlation coefficients for high performers' work session 103

3.19 Descriptive statistics for low performers (values are in mean (standard deviation)) 104

3.20 Mixed factors ANOVA results for low performers' adjustment session............ 105

3.21 Correlation coefficients for adjustment session for low performers................... 106

3.22 Mixed factors ANOVA results for low performers' work session 107

3.23 Tukey's post hoc comparisons for low performers' work session 107

3.24 Correlation coefficients for work session 108

4.1 Mixed factors ANOVA results for the overall test session................................. 151

4.2 Mixed factor ANOVA results for independent performance measures in each scenario for both driver groups 152

4.3 Tukey's post hoc comparisons for interactions between experience and gender for overall stage and independent events in each scenario 153

4.4 Descriptive statistics for the significantly different performance....................... 155

4.5 Descriptive statistics for the dependent variables for the experienced drivers for overall (values are in mean (standard deviation)) 157

4.6 Mixed factor ANOVA results for the experienced drivers for overall session... 158

4.7 Tukey's post hoc results for experienced drivers overall session for stage 159

4.8 Tukey's post hoc results for the experience by gender interaction in the overall session for experienced drivers 162

4.9 Correlation coefficients for the overall session for experienced drivers 163

4.10 Mixed factor measures ANOVA results for scenario specific performance measures for experienced drivers' city scenario 165

4.11 Descriptive statistics for gender difference for experienced drivers' city scenario (values are in mean (standard deviation)).. 166

4.12 Tukey's post hoc results for experienced drivers' city scenario 167

4.13 Tukey's post hoc results for the gender by stage interaction effect for experienced drivers' city scenario 169

4.14 Correlation coefficient for experienced drivers for city scenario 170

4.15 Mixed factor ANOVA results for experienced drivers' highway scenario 171

4.16 Tukey's post hoc results for experienced drivers' highway scenario 172

4.17 Tukey's post hoc results for the gender by stage interaction effect for experienced drivers' highway scenario... 173

4.18 Correlation coefficients for experienced drivers' highway scenario 174

4.19 Mixed factors ANOVA results for experienced drivers' rural scenario 175

4.20 Tukey's post hoc results for experienced drivers' rural scenario 176

4.21 Tukey's post hoc results for the gender by stage interaction effect for experienced drivers' rural scenario 178

4.22 Correlation coefficients for experienced drivers' rural scenario 179

4.23 Descriptive statistics for novice drivers overall session (values are in mean (standard deviation)) 180

4.24 Mixed factor ANOVA results for novice drivers overall session 182

4.25 Tukey's post hoc results for novice drivers overall session 183

4.26 Tukey's post hoc results for the gender by stage interaction effect for novice driver overall session 185

4.27 Correlation coefficients for novice drivers overall session................................ 186

4.28 Mixed factors ANOVA results for novice drivers' city scenario 188

4.29 Tukey's post hoc results for novice drivers' city scenario stages....................... 189

4.30 Tukey's post hoc results for the gender by stage interaction effects for novice drivers' city scenario... 191

4.31 Correlation coefficients for novice drivers' city scenario.................................. 192

4.32 Mixed factor ANOVA results for novice drivers' highway scenario 193

4.33 Tukey's post hoc results for novice drivers' highway scenario stages 194

4.34 Tukey's post hoc results for the gender by stage interaction effect for novice drivers' highway scenario 196

4.35 Correlation coefficients for novice drivers' highway scenario.......................... 197

4.36 Mixed factors ANOVA results for novice drivers' rural scenario...................... 198

4.37 Tukey's post hoc results for novice drivers' rural scenario stages 199

4.38 Tukey's post hoc results for the gender by stage interaction effect for novice drivers' rural scenario 201

4.39 Correlation coefficients for novice drivers' rural scenario 202

4.40 Comparisons between experienced and novice drivers for perception time....... 204

4.41 Comparisons between experienced and novice drivers for movement time....... 206

4.42 Comparisons between experienced and novice drivers for total brake time 208

4.43 Comparisons between experienced and novice drivers for collision rate........... 209

4.44 Comparisons between experienced and novice drivers for lost control rates..... 211

4.45 Comparisons between experienced and novice drivers for parking duration 213

4.46 Comparisons between experienced and novice drivers for parking failure rate . 214

4.47 Comparisons between experienced and novice drivers for lane deviation (highway scenario) 216

4.48 Comparisons between experienced and novice drivers for lane deviation in rural scenario 218

4.49 Comparisons between experienced and novice drivers for S-curve performance 219

4.50 Comparison between experienced and novice drivers for subjective mental workload measures slopes and thermography 221

4.51 Correlation coefficients between performance, subjective workload meausres, and thermography for novice drivers 233

4.52 Correlation coefficients between subjective workload measures and thermography for novice drivers 234

4.53 Sufficient training time for each performance 235

4.54 Adapting time for each performance for experienced drivers 235

4.55 Mean for gender in each variables for experienced and novice drivers 238

LIST OF FIGURES

1.1 Performance and mental workload 4

2.1 Mental workload factors 16

2.2 Thermal Image with ROI 29

2.3 Trends of ΔNT and ΔFT 35

2.4 Trends for each key question in each block on average 377

2.5 Trends of the MCH and SWAT 39

2.6 Trends of accuracy and RT 41

2.7 Trends of ΔNT and subjective workload measures 43

2.8 Trends of ΔNT and performance measures 44

2.9 A participant's thermal readings in each block (Accuracy rate %) 48

2.10 Trend of accuracy and nose temperature standard deviations across task blocks 49

2.11 Workload, performance, and task demands relationship (adapted from de Waard, 1996) 54

3.1 Rasmussen's taxonomy and Wickens et al task performance hierarchy 69

3.2 Relationship between performance and mental workload 70

3.3 Relationship mental workload and performance based on Meister's model (1976) 71

3.4 Relationship between operator performance and time constraints (Johannsen, 1979). 73

3.5 Speed accuracy trade-off 74

3.6 Thermal Image with ROI 84

3.7 Thermography trend of adjustment session for high performers 114

3.8 Thermography trend of evaluation session in low performers 115

3.9 ΔNT Standard Deviation trend for the performance groups 118

3.10 Trend of ΔNT in high and low performers 119

4.1 Comparison of cumulative system costs with and without simulation (Harrell et al., 2004). 136

4.2 Thermal Image with ROI 144

4.3 Hierarchy diagram for analysis 148

4.4 Interaction effect between experience and gender on parking duration 154

4.5 Interaction effect between stage and gender on ΔFT 160

4.6 Interaction effect between stage and gender on MCH ratings 161

4.7 Stage by gender interaction effect in ΔFT for experienced drivers'city scenario 168

4.8 Gender by stage interaction effect in ΔFT for experienced drivers' highway scenario 172

4.9 Gender by stage interaction effect in ΔFT for experienced drivers' rural scenario 177

4.10 Gender by stage interaction effect in ΔFT for novice drivers overall session 184

4.11 Gender by stage interaction effect for ΔFT novice drivers' city scenario stages 190

4.12 Gender by stage interaction effect in ΔFT for novice drivers' highway scenario stages 195

4.13 Gender by stage interaction effect in ΔFT for novice drivers' rural scenario 200

4.14 Trends of perception time for experienced and novice drivers 203

4.15 Trends of movement time for experienced and novice drivers 205

4.16 Trends of total brake time for experienced and novice drivers 207

4.17 Trends of collision for experienced and novice drivers 209

4.18 Trends of lost control for experienced and novice drivers 211

4.19 Trends of parking duration for experienced and novice drivers 212

4.20 Trends of parking failure for experienced and novice drivers (average number of occurrences) 214

4.21 Trends of lane deviation for experienced and novice drivers for lane deviation for highway scenario (average number of occurrences) 215

4.22 Trends of lane deviation for experienced and novice drivers in rural scenario (average number of occurrences) 217

4.23 Trends of lane deviation for experienced and novice drivers (average failed number) 219

4.24 SWAT rating trends for experienced and novice drivers 220

4.25 MCH rating trends for experienced and novice drivers 221

4.26 Trend of the ΔNT for experienced and novice drivers for overall stages 222

4.27 Fatigue level fluctuations by time (minutes) (Nilsson, 1997) 229

4.28 Trend of the drivers' performances that was found no differences between experienced and novice drives. 230

CHAPTER 1

INTRODUCTION

Mental workload is a complex and ambiguous concept (Jex, 1988; Reid and Nygren, 1988). Several definitions of mental workload currently exist, and no one universal definition is widely accepted. The Federal Aviation Administration (FAA, 2005) defined mental workload as "…the physiological and mental demands that occur while performing a task or combination of tasks." Hart and Staveland (1988) add to this definition that "…workload is not an inherent property, but rather it emerges from the interaction between the requirements of a task, the circumstances under which it is performed, and the skills, behaviors, and perceptions of the operator." Also, Hancock and Chignell (1988) concluded that mental workload has an impact on human task performance. They indicate that "successful performance depends upon the reconciliation of the complexity and difficulty of the imposed task with the time within which the goal must be achieved".

- The complexity of the mental workload concept and the lack of a consistent definition have impacted the development of effective mental workload measurement tools. Four general classifications of mental workload measurement methods exist: primary task measures, secondary task measures, subjective measures, and physiological measures. Current research trends are to use objective measures to quantify workload; therefore, increased emphasis is being placed on physiological measures. However, there are limitations that affect the applicability and utility of these measures.

Researchers have been investigating the use of novel methods for quantifying mental workload. Physiological measures present a promising area of research as they are not easily influenced by the human operator. Many current physiological measures require sensors or other equipment to be attached to the human operator, which may interfere with an individual's ability to perform the work (e.g., limiting motion) and increase stress levels, thereby artificially increasing mental workload levels. Therefore, the use of non-contact physiological techniques is being investigated.

Thermography has been identified as a potential non-contact physiological mental workload measure as it can directly measure the physiological responses of the autonomic nervous system (ANS) believed to be impacted by mental workload levels (Veltman and Vos, 2005; Kang et al., 2006; Boucsein, 1993; Genno et al., 1997[a,b]). Thermography provides a measure of an object's surface area temperature (in this case the human). Changes in facial temperatures have been found to be affected by mental workload levels (Veltman and Vos, 2005; Kang et al., 2006), though further studies are needed to validate the use of thermography to quantify mental workload.

Effective mental workload measurement is critical as researchers have found mental workload to significantly affect human performance, particularly in complex systems (Gopher and Donchin, 1986; Colle et al, 1988; Hancock and Meshkati, 1988; Moray, 1979; O'Donnell and Eggemeier, 1986; Jex, 1988). Both low and high mental workload levels can degrade performance (Lysaght et al., 1989) and there is an optimal level of mental workload that corresponds to "optimal" performance levels (e.g. Young and Stanton, 2001; Kaber et al., 2001). The relationship between mental workload and performance is consistent with the Yerkes-Dodson's law (1908), which predicts a parabolic function between these parameters (Figure 1.1). Moray (1988) pointed out that optimizing mental workload levels could reduce human error, improve system safety, increase productivity, and improve operator satisfaction, though no well documented strategies currently exist to quantify the optimal workload level.

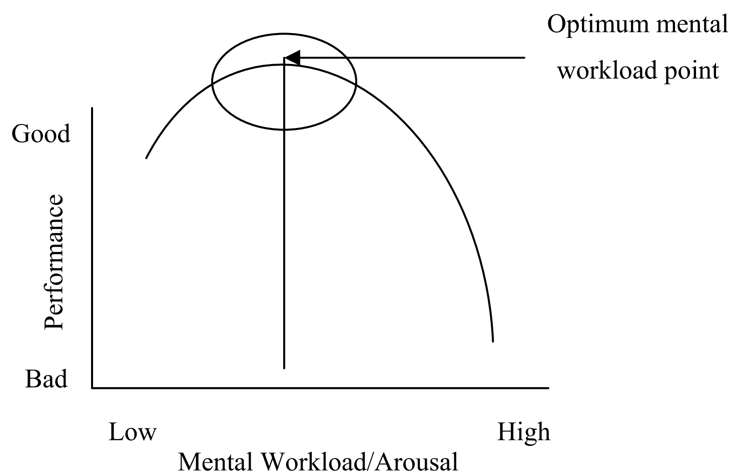


Figure 1.1

Performance and mental workload

In addition to optimizing human performance, mental workload has been linked to learning and training. Learning is defined as "the acquisition of knowledge of skills through experience, practice, or study, or by being taught" (New Oxford American Dictionary, 2001). Salas et al. (2006) defined training as, "the systematic acquisition of knowledge (i.e., what we need to know), skills (i.e., what we need to do), and attitudes (i.e., what we need to feel) that together lead to improved performance in a particular environment". Therefore, learning is achieved through training and training provides the potential to decrease cognitive resource and workload levels (Tsang and Vidulich, 2006). Jex (1988) indicated that training and practice contribute as psychophysiological influences in reducing a person's mental workload. Thus, mental workload can be reduced when operators acquire a required skill or knowledge through training.

The American Society for Training and Development's "State of the Industry Report" (ASTD, 2003) states that employee training costs are increasing. It is estimated that training costs increased from $734 in 2001 to $826 in 2003 per employee and employees spent an estimated 24 hours in 2001 to 28 hours in 2002 in training. These costs result in an estimated $54.2 – $200 billion being spent each year in training and developing employees' skills (Salas and Cannon-Bowers, 2001; Galvin, 2002).

Given these numbers, techniques that can identify when sufficient training has occurred would reduce not only the time spent on training, but also reduce the costs associated with training. Sufficient training times are not well documented in the literature. In fact, training times are typically preset prior to the trainee entering the training environment. As individual differences are known to affect learning (Sadler-Smith and Smith, 2004; Chen and Macredie, 2004) it is reasonable to assume that training times would also differ across individuals.

1.1 PROBLEM STATEMENT

No single mental workload assessment tool is widely accepted, creating the need for continued research into new mental workload measurement tools. Current trends in mental workload research on non-contact assessment measures means the role of non-contact physiological and primary tasks measures are gaining in importance. However, traditional physiological workload tools cannot be classified as non-contact methods. Also, as mental workload has documented affects on human performance, and there is an optimal mental workload level that maximizes performance (Kaber et al, 2001; Young & Stanton, 2001; Moray, 1988; Johannsen, 1979; Meister; 1976), methods for determining optimal workload levels are needed, and current methods are not well defined or do not exist.

Further, training is a commonly used method for skill acquisition in the workplace. Learning, a goal of training, may increase mental workload until the skill or knowledge is retained (Jex, 1988; Baddeley, 1986). However, few studies exist that illustrate the relationship between mental workload and learning. If indeed a relationship does exist, then the use of non-contact physiological mental workload assessment techniques could be used to assess learning and prescribe suggested training times to maximize learning, while minimizing training costs on organizations.

1.2 OBJECTIVES

The objectives of this study are to: (1) investigate the efficacy of using thermography as a non-contact physiological measure to quantify mental workload, (2) quantify and describe the relationship between mental workload and learning/training, and (3) introduce a method to determine a sufficient training time and an optimal human performance level for a novel task by using thermography.

The specific research questions to be addressed are:

1. Can thermography measure mental workload?
2. Is there a relationship between mental workload and learning process (training)?
3. Can thermal facial readings be used to identify sufficient training times associated with learning a new task?
4. Can optimal performance and workload levels be determined by thermal facial readings?

1.3 RESEARCH OUTLINE

Three studies were conducted to answer the above research questions.

Study 1: This study was conducted to quantify the relationship between thermal readings, subjective workload assessment ratings, such as Modified Cooper-Harper or SWAT, and individual performance for a novel task. Preliminary evidence currently exists to support significant linear relationship between nose temperature readings, workload, and learning (Kang et al., 2006).

Study 2: This study was an investigation into a potential methodology for determining the workload level that maximizes performance levels. The methodology was based on psychophysical techniques that are based on the premise that individuals can identify the optimal levels of work demands over a shortened period of time. This study also quantified the relationship between thermal readings, subjective workload assessment ratings, and individual performance for a novel task, and provided supports for study 1 findings.

Study 3: The third study was conducted using a driving simulator to evaluate efficacy of thermography assessments of the workload obtained during a practical driving application. Thermal readings were used to identify relationships between mental workload and training time. Also, driver's performance data and subjective workload assessment ratings were collected to identify and quantify the relationship with thermal readings in pursuit of training process. This study was a preliminary investigation of using thermography to define training time.

The expected results of this research are the initial validation of using thermography to quantify mental workload, introduction of a methodology to define optimal mental workload levels to maximize performance, and preliminary support for the use of thermography to define training times. These contributions are novel or provide support for newer techniques in field of mental workload measurement. Potential implications from the findings of this research are improved safety, increased productivity, reduced human error, and reduced training time and cost.

1.4 REFERENCES

Andre, A.D., & Wickens, C.D. (1995). When users want what's not best for them. *Ergonomics in Design:The Quarterly of Human Factors Applications 3(4)*, Oct. 10-14(5).

Alvager, T., Anderson, E., French, V.A., Putman, G., & Shi, L. (1999). An artificial neural network study of the relationship between arousal, task difficulty and learning, *IJCNN '99. International Joint Conference on Neural Networks*, 5. 3618 – 3621.

American Society for Training and Development (ASTD) (2003). *ASTD 2003 state of the Industry Report Executive Summary,* Retrieved June 3, 2004, From http://www.astd.org/NR/rdonlyres/6EBE2E82-1D29-48A7-8A3A-357649BB6DB6/0/SOIR_2003_Executive_Summary.pdf.

Baddeley, A.D. (1986). *Working memory.* Oxford, UK: Oxford University Press.

Casali, J.G., & Wierwille, W.W. (1983). A comparison of rating scale, secondary task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors*, 25, 623-642.

Chen, S.Y., & Macredie, R.D. (2004). Cognitive modeling of student learning in web-based instructional programs. *International Journal of Human-Computer Interaction*, 17, 375-402.

Colle, H.A., Amell, J.R., Ewry, M.E., & Jenkins, M.L. (1988). Capacity equivalence curves: a double trade-off curve method for equating task performance. *Human Factors*, 30, 645-656.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games, *Applied Artificial intelligence*, 16, 555-575.

Derrick, W.L. (1988). Dimensions of operator workload, *Human Factors*, 30(1): 95-110.

Federal Aviation Administration (FAA). (2005). *FAA Human Factors Awareness Course*. Retrieved December 7, 2005. From http://wfnetwork.bc.edu/glossary_entry.php?term=Workload,%20Definition(s)%20of&area=All

Galvin, T. (2002). 2002 Industry Report, *Training*, 39, October 24-37.

Gaillard, A.W.K. (1993). Comparing the concepts of mental load and stress. *Ergonomics*, 36, 991-1005.

Gawron, V.J. (2000).*Human Performance Measures Handbook*. Mahwah, NJ: Lawrence Erlbaum Associates..

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. BoŠ, L. Kaufman, & J. Thomas (Eds.). *Handbook of Perception and Human Performance : Volume II. Cognitive Processes and Performance* (pp. 41-49). New York, NJ: Wiley.

Hancock, P.A., & Chignell, M.H. (1988). Mental workload dynamics in adaptive interface design. *IEEE Transactions on Systems, Man and Cybernetics*, 18, 647-658.

Hancock, P. A., & Meshkati, N. (1988). *Human Mental Workload*. Elsevier Science: Amsterdam.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Handcock & N. Meshkati, (Eds.),*Human Mental Workload*(pp. 139-178).Amsterdam: North-Holland.

Hockey, G.R.J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of Perception and Human Performance*, Vol. II 44-1- 44-49.

Jex, H. R. (1988). Measuring mental workload: Problems, progress, and promises, In P.A. Hancock & N.Meshkati, (Eds.),*Human Mental Workload* (pp. 5-39). New York: North-Holland.

Johannsen, G. (1979). Workload and workload measurement. In N, Moray (Ed.) *Mental Workload: Its Theory and Measurement* (pp. 3-12). New York:Plenum.

Kaber, D. B., Riley, J. M., Tan K., & Endsley, M. R. (2001). On the Design of Adaptive Automation for Complex Systems.*International Journal of Cognitive Ergonomics*, 5, 37 – 57.

Kang, J., McGinley, J., McFadyen, G., & Babski-Reeves, K. (2006). Determining learning level and effective training times using thermograhy. *Proceedings of 25thArmy Science Conference*, Orlando, FL.

Kramer, A.F. (1987). Special issue on cognitive psychophysiology. *Human Factors*, 9, 129-127.

Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D. Linton, P.M., Wierwille, W.W., Zaklad, A.L., Bittner, A.C., & Wherry, R.J. (1989). *Operator workload: comprehensive review and evaluation of operator workload methodologies* (Technical Report 851). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Meister, D. (1976). *Behavioral Foundations of System Development*. New York: Wiley.

Moray, N. (1979). *Mental Workload: Its Theory and Measurement*.New York: Plenum.

Moray, N. (1988). Mental workload since 1979. In D.J. Oborne (Ed.), *International Reviews of Ergonomics*: Vol. 2.*Current trends in human factors research and pratice (International reviews of ergonomics)*(pp. 123-150).London: Taylor & Francis.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. BoŠ , L. Kaufman, & J. Thomas (Eds.), *Handbook of Perception and Human Performance : Volume II.Cognitive Processes and Performance* (pp. 42.1-42.9).New York: Wiley.

Reid, G.B. & Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock & N. Meshkati. (Eds). *Human Mental Workload* (pp. 185-213).Amsterdam: North-Holland.

Saldler-Smith, E., & Smith, P.J. (2004). Strategies for accommodating individuals' styles and preferences in flexible learning programmes. *British Journal of Educational Technology*, 35, 395-412.

Sales, E., & Cannon-Bowers, J.A. (2001). The Science of Training: A Decade of Progress. *Annual Review of Psychology*, 52, 471-499.

Salas, E., Wilson, K.A., Priest, H.A., & Guthrie, J.W. (2006). Design, delivery, and evaluation of training systems, In G. Salvendy (Ed.). *Handbook of Human Factors and Ergonomics* (pp. 234-268). New Jersey: John Wiley& Sons.

Sanders, M.S., & McCormick E., J. (1993). *Human Factors in Engineering and Design*. New York, McGRAW-Hill.

Sheridan, T. (1980). Mental workload: what is it? Why bother with it?. *Human Factors Society Bulletin*, 23, 1-2.

Slocum, G.K., Williges, B.H., & Roscoe, S.N. (1971). Meaningful shape coding for aircraft switch knobs. *Aviation Research Monographs,* 1, 27-40.

Swezey, R.W., & Pearlsten, R.B. (2001). Selection, training, and development of personnel. In G. Salvendy (Ed.). *Handbook of Industrial Engineering: Technology and Operations Management(3rd ed.)* (pp. 920-947). New York: John Wiley & Sons.

Tsang, P.S., & Vidulich M.A. (2006). Mental workload and situation awareness. In G. Salvendy (Ed.). *Handbook of Human Factors and Ergonomics (3rd ed.)* (pp. 234-268). New Jersey: John Wiley & Sons.

Thompson, L.A., Williams, k.l., L'Esperance, P.R., & Cornelius, J. (2001). Context-dependent memory under stressful conditions: The case of skydiving, *Human Factors,* 43, 611-619.

Wickens, C.D. (1979). Measures of workload, stress and secondary tasks. In N. Moray (Ed.), *Mental Workload: Its Theory and Measurement*(pp. 79-99). New York: Plenum.

Wickens, C.D., Gordon, S.E., & Liu, Y. (1998). *An Introduction to Human Factors Engineering.*New York: Addison Wesley Longman, Inc.

Wickens, C.D., & Hollands, J.G. (2000). Engineering Psychology and Human Performance*(3rd ed.).* New Jersey: Prentice_Hall.

Wierwille, W.W., & Casali, J.G. (1983). A validated rating scale for global mental workload measurement applications.*Proceedings of the Human Factors Society 27th Annual Meeting*, Norfolk, VA, 129-133.

Yerkes, R.M. & Dodson, J.D., (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.

Young, M.S. & Stanton, N.A. (2001). Mental workload: theory, measurement, and application. In W. Karwowski(Ed.),*International Encyclopedia of Ergonomics and Human Factors* (pp. 507-509).New York: Taylor and Francis.

CHAPTER 2

EFFICACY OF USING THERMOGRAPHY TO QUANTIFYING MENTALWORKLOAD WHILE PERFORMING A NOVEL COGNITIVE TASK

2.1 ABSTRACT

Objective: The objective of this study was to investigate the efficacy of using thermography to quantify the relationship between mental workload and facial temperature changes during learning of a novel task.

Methods: Twenty participants, 10 male and 10 female, completed 7 blocks of an alpha-numeric task. Changes in nose and forehead temperature, task accuracy, reaction time, and two subjective mental workload ratings (MCH ratings and SWAT) were collected.

Results: Strong and significant correlations were found among thermography (ΔNT), performance, subjective workload ratings. Sufficient training time for this task, as identified through analysis of thermal readings is described as training (1 block) and skill mastery (2 blocks). This finding was supported by performance measures and subjective workload ratings.

Conclusion: Thermography is a valid, objective, non-invasive mental workload measure and is more practical than current subjective mental workload measures. Thermography is also capable of identifying sufficient training time for cognitive tasks.

2.2 INTRODUCTION

Learning is required to diminish the distinction between the current performance level of an operator and task demands of a specific job. During the learning process, however, mental workload is imposed on the operator/learner until they acquire required skill(s) and knowledge to perform a specific job (Jex, 1988).Researchers suggest that the concept of mental workload is defined in terms of the human's limited processing resources, such as the rate at which decisions are made, and the difficulty of making the decisions (Wickens, 1979; O'Donnell and Eggemeier, 1986; Moray, 1979).Mental workload can be described as the operator's perceived gap between the operator's current performance ability and the current task demands (e.g. from learning a new task). Researchers have indicated that operators' (as learners) mental workload can be reduced by training and practice (Moray, 1988; Jex, 1988; Wickens and Hollands, 2000).

Several definitions for mental workload currently exist. The Federal Aviation Administration (FAA, 2005) defined workload as, "the physiological and mental demands that occur while performing a task or a combination of tasks." Hart and Staveland (1988) modified the definition to state that mental workload is variable and is affected by the human-system interaction and operators' individual differences. For example, research has shown that operators performing the same task can report differing workload levels because of individual differences in mental ability, motivation, training, effort, or physical capabilities (Moray, 1988; Tsang and Vidulich, 2006). In the current study, mental workload is defined as the quantity of information that can be processed at a specific time at a specific level of difficulty.

Regardless of the definition, mental workload levels are affected directly and indirectly by many factors (figure 2.1); such as stress (Gaillard, 1993), personnel condition (skill or ability), attention, and fatigue (Wickens and Hollands, 2000). Current definitions suggest that mental workload affects human performance and psychophysiological activity. Further, it is assumed that changes in mental workload result in observable and measurable physiological changes. These changes can be used to develop alternative workload measurement assessment techniques.

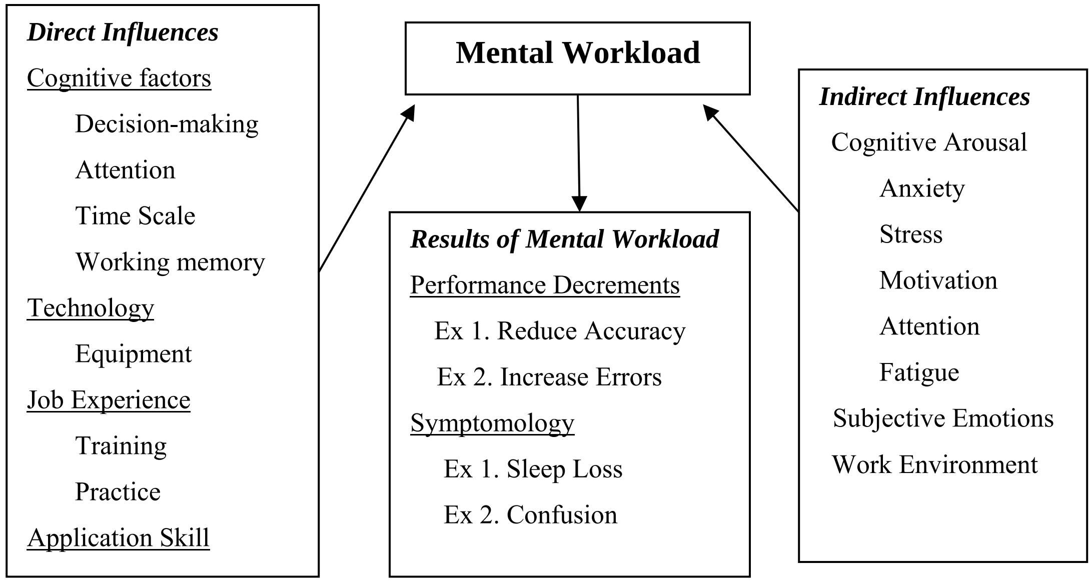


Figure 2.1

Mental workload factors

2.3 BACKGROUND

2.3.1 MENTAL WORKLOAD MEASUREMENT

Various techniques for measuring mental workload exist and can be categorized into primary task measures, secondary task measures, physiological measures, or subjective measures (Table 2.1). For each measurement category, there are issues with data collection and the establishment of a relationship to workload (Table 2.2).

Primary task measures are associated with evaluating performance measures on the major task of the operators, such as typing speed, yaw and pitch for airplanes, or learning comprehension with a particular method of instruction (Wickens and Hollands, 2000). As cognitive demands of a task change, changes in operator performance can be detected by primary task measures (Tsang and Vidulich, 2006). The primary problem with using primary task measures for mental workload measurement is that they are task-specific, making it difficult to compare between different tasks (Sanders and McCormick, 1993).

Table 2.1

Workload measurement methods

Measurement	Description	Example
Primary task measures	Primary task measures evaluate the most directly related task performed on the system or operator such as computer data-entry speed, driving deviations from the center of the land, or learning comprehension with a particular method of instruction (Wickens and Hollands, 2000).	Workload defined as the time required to perform a task divided by the time available to perform the task, such as duration and frequency (e.g. Wierwille and Connor, 1983; frequency)
Secondary task measures	These measures ask an operator to perform the primary task as major task with a secondary task as minor task. It induces the operator to use their spare attention or capacity to perform a secondary task (task's specified requirement) (Gawron, 2000).	Driving secondary task (e.g.Johnson and Haygood, 1984; visual choice with varying the road width.)

Table 2.1 Continued

Subjective task measures	Subjective measures quantify mental workload with rating workload on a subjective scale. The rating relies on subjective perception base on an operator's actual experience (Sheridan, 1980; Wickens et al, 1998).	NASA Task Load Index scale (Hart and Staveland, 1988), the Subjective Workload Assessment Technique (Reid and Nygren, 1988), present a modified Cooper-Harper rating scale (Wierwille and Casali,1983)
Physiological task measures	Physiological measures quantify mental workload with a single-resource model of information processing based on autonomic or central nervous system activity (Sanders and McCornick, 1993; Kramer, 1987).	Heart rate (e.g. Hankins and Wilson, 1998), Respiratory measure (e.g. Wilson, 1992), EEGs (e.g. Gevins and Smith, 2003).

Table 2.2

Mental workload measurement issues

Measurement	Data collection issues
Primary task measures	A problem with primary mental workload measures in that they are task-specific, making it difficult to compare between different tasks (Sanders and McCormick, 1993).
Secondary task measures	Participant deals with secondary task as primary task during the task performance measure if researcher chooses the secondary task incorrectly (Sanders and McCormick, 1993).
Subject task measures	Subjective measures also have the limitation that people's subjective perception cannot always coincide with their task performance (Andre and Wickens, 1995) because subjective perception can be affected by many factors such as operator's emotion, fatigue, and etc.
Physiological task measures	Physiological measures can provide objective information of a subject's mental workload, though they may impose limitations on task performance as well as physical discomfort and contact stress (Kataoka et al, 1998; Genno et al., 1997a).

Secondary task measures are some of the most widely used mental workload measures. These measures ask an operator to perform a task in addition to their primary task, thereby requiring operators to allocate spare capacities or attentional resources to complete the secondary task (Gawron, 2000). If performance on the secondary task requires higher mental workload, there are fewer mental resources available for the completion of the secondarily. Secondary task measures are more sensitive in measuring mental workload than primary task measures because they are believed to demonstrate difficulty level differences between primary tasks (Wickens et al, 1998; Slocum et al., 1971; Gawron, 2000). However, it may be infeasible to impose a secondary task due to the criticality of the primary task (driving, flying, emergency medical technician, etc.). Therefore, the applicability and utility of these measures are limited.

Subjective measures ask operators to rate their workload, typically on a scale, based on their subjective perceptions of their experience (Sheridan, 1980; Wickens et al, 1998). The advantages of these methods are they are easy to administer and to obtain ratings (Sanders and McCormick, 1993; Casali and Wierwille, 1983). Some measures elicit a unidimensional rating of mental workload (e.g., the Modified Cooper Harper Scale, Wierwille and Casali, 1983), whereas others combine ratings along multiple dimensions (e.g., the NASA Task Load Index, Hart and Staveland, 1988; or Subjective Workload Assessment Technique, Reid and Nygren, 1988), The limitation of subjective workload measures are that operators perceptions of mental workload do not always coincide with task performance (Andre and Wickens, 1995). Further, mental workload ratings can be influenced by other factors not related to the task, such as emotional stress, fatigue, etc. (Gaillard, 1993; Wickens et al., 1998). It also is difficult to distinguish external task demand difficulty from actual workload if the tool questions or scales are not well defined (O'Donnell and Eggemeier, 1986). Though some consider subjective mental workload measures to be the most accurate measure of mental workload, their primary limitation is that they intrude upon task performance. Operators must stop their work to complete the rating process, and therefore, may also be limited in terms of applicability and utility.

Physiological measures quantify mental workload with a single-resource model of information processing based on autonomic or central nervous system activity (Sanders and McCornick, 1993; Kramer, 1987; Tsang and Vidulich, 2006). The central nervous system (CNS) includes the brain, brain stem, and spinal cord cell, and CNS measures are used to detect brain activity. Activities for the CNS can be autonomic (such as heart rate changes and blood vessel constriction/dilation) or voluntary (such as muscle contractions). It is the autonomic responses that are of most interest in mental workload measurement as these are physiological responses that are not controlled or influenced by conscious activities. The autonomic nervous system (ANS) is divided into the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS). SNS provides extra activation to the body in emergency situations (stress state) involving a fight or flight reaction while PNS helps to maintain homeostasis limits within the body system by relaxing the body as a regulatory system. Mental stress and emotional state are strong triggers to activate the SNS. SNS stimulation increases mental activity, heart rate, and pupil size. It also contracts the smooth muscle of the organs that constricts blood vessels and pores in the skin. Vasoconstriction is related to decreases skin surface temperature due to decreased blood flow in tissues. On the other hand, PNS leads to decreased heart rate and pupil size, but it has no effect on mental activity, muscle, or skin (Guyton and Hall, 2006).

Physiological measures are often continuous measures (rather than discrete as is the case for other mental workload measures) and can capture manifestations of mental workload unobtrusively, unlike other measures (e.g., secondary and subjective measures) (Kramer, 1987). Physiological measures that have been used to quantify mental workload include heart rate variability, evoked brain potentials, respiration rate, and body fluid chemistry. However, it is uncertain how physiological measures relate to all dimensions of mental workload (Moray, 1988). Further, the use of equipment to record these responses may alter operator behaviors, restrict their mobility, or introduce stress, thereby increasing mental workload level artificially (Kataoka et al, 1998; Genno et al., 1997a).Therefore, physiological measures usually are used with other mental workload measures (Kramer, 1987).

Despite the wide range of mental workload measurement techniques, no single measure has been widely accepted. Primary task measures are not really a mental workload measure by themselves, as they focus on operator performance, and a lack of secondary task performance decrement may only indicate that participants wrongly treat the secondary task as the primary task. Of available measures for considering mental workload, subjective measures most nearly capture the mental workload concept (Sheridan, 1980) because mental workload is generally accepted as a subjective perception. Subjective mental workload measures have been used with the rationalization that many subjective factors can affect mental workload directly or indirectly (Sheridan, 1980; Gaillard, 1993; Xie and Salvendy, 2000). However, these tools require tasks to be interrupted for data collection or rely on operator recall to estimate workload levels making them unreliable. Physiological measures provide "continuous and objective information about the state of an operator" (Veltman and Vos, 2005), and can reflect the amount of mental effort that an operator has to invest in order to perform the task adequately. However, physiological measures may impose limitations on task performance (e.g., limit range of motion, introduce distraction, etc.) as well as physical discomfort and contact stress (Kataoka et al, 1998; Genno et al., 1997a).Therefore, there is a need to develop alternative objective measures of mental workload that do not require equipment contacts with the operators.

2.3.2 THERMOGRAHY AS A MENTAL WORKLOAD MEASURE

Researchers have shown that sensations such as stress, anxiety or fatigue bring about considerable levels of change in body temperatures (Genno et al, 1997a). Observations from previous studies have shown that these temperature changes are significantly observable in the facial areas due to increased autonomic nerve activity associated with sensations (e.g., Drummond et al., 2003; Naemura et al, 1993; Boucsein, 1993; Vernet-Maury et al, 1993, Genno et al.,1997[a,b]). Genno et al., (1997[a,b]) used thermography in their studies and demonstrated that a change in nose temperature can be used to indicate mental fatigue. In fact they found that people facing sudden anxious situations experience an immediate temperature decrease in the nose area.

Facial skin temperature, in particular, is an indicator of human sensation levels because the blood flow in the face fluctuates with the ANS (Genno et al, 1997[a,b], Kataoka et al, 1998). It has been shown that there is a high correlation between mental stress (a concept related to workload) and nose temperature, where nose temperature decreases with an appearance of a stressful noise (Genno et al, 1997[b]; Naemura et al, 1993; Kataoka et al., 1998; Veltoman and Vos, 2005; Kang et al, 2006), while forehead temperature is stable in the presence or absence of mental stress (Genno et al, 1997[b]; Stoll 1964; Kataoka et al., 1998; Veltoman and Vos, 2005)

Researchers have indicated that skin temperature changes may be a function of mental workload (Green and Shellenberger, 1991; Trujillo, 1998). Trujillo (1998) found that the dorsal surface temperature of the index finger of pilots was lower when pilots reported higher subjective workload ratings during simulated flight. Facial skin temperature can provide a physiological automatic measurement without placing physical contact stress on the operator as other physiological measures (Genno et al, 1997[a]). Veltman and Vos (2005) demonstrated the relationship between mental workload and a change in facial temperature. However, validation of thermography as an objective workload measurement technique is limited.

2.4 OBJECTIVES

The objective of this study was to assess the efficacy of using thermography to quantify mental workload. This research quantified the relationship between facial temperature changes and the human learning process. Human learning process was assessed using performance measures, particularly accuracy and reaction time. This study also served as the basis for a more extensive investigation of using thermography to identify when learning is occurring and sufficient training times.

2.5 HYPOTHESES

1. Nose temperature will be affected by mental workload manipulations. However, forehead temperature remains mostly constant.
2. Nose temperature will be correlated with subjective measures of mental workload.
3. Performance will be correlated with mental workload, specifically nose temperature readings.
4. Performance will have an asymptotic relationship with time (i.e., learning).
5. Mental workload will decrease with time (i.e., learning).
6. No gender differences will be found.

2.6 METHODOLOGY

2.6.1 EXPERIMENTAL DESIGN

A one factor within subject design was used to assess learning (block) on thermographic readings of the face, subjective workload assessment ratings, and performance. A novel, alpha-numeric task was used in this study consisting of 7 test blocks. Mental workload was measured using thermal imaging of the face and subjective workload assessment ratings. Exposures to test trials were determined using a Balanced Latin Square.

2.6.2 PARTICIPANTS

Twenty participants, 10 males and 10 females, completed experimental protocols. Participants ranged in age from 19 to 26 years. Participants were recruited from the Mississippi State University student community. No inclusion or exclusion criteria were used to screen participants.

Estimation of sample size

Sample size estimates were based on data by Kang et al (2006), which used a similar methodology as the one proposed below. Power analysis was performed to estimate a sample size that is large enough to detect differences between paired data from this study. In that study, data pre- and post-learning was used in the test statistic. The general test statistic for paired data would be the two-side t-test. Sample size formula for this study is given by (Montgomery and Runger, 2003).

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2 (\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2}$$

Hypothesis H_0: $\mu_1 - \mu_2 = \Delta_0$

Δ is a given difference between two means.

σ is the standard deviation for two population

Sample sizes ranging from 20 to 33 participants were obtained for various levels of β (0.20. 0.15, 0.10, 0.05) and $\alpha = 0.05$. As this study was exploratory, it was determined that an acceptable level of power was 0.80, resulting in a sample size of 20.

2.6.3 TASK DESCRIPTION

This study employed the alphabet arithmetic task used by Logan and Klapp (1991) to assess learning a novel task. In the alphabet arithmetic task, subjects verify equations of the form C+2=? (E), D+3= ? (G), and E+4=? (I). Subjects were asked to compute problems by adding the letters C, D, or E with a numbers 1 through 4.

2.6.4 INDEPENDENT VARIABLES

Blocks and gender were treated as independent variables. The experiment consisted of 7 task blocks, 6 minutes and 24 seconds each in length, with a 3 minute rest period between blocks. Each block consisted of 8 question sets that consisted of 12 questions (one block = 96 questions). The same 12 questions were used in each set, though the questions were randomly ordered, excluding a single question. A key question (D+4=?) appeared in the 8th place of each set allowing for an assessment at a consistent point in time across task blocks. Participants were provided with a 4-second time limit to respond to each question, potentially imposing time stress on participants.

2.6.5 DEPENDENT VARIABLES

Dependent variables for this study included facial thermal readings, two subjective workload ratings, and performance measures. Thermal readings were collected continuously during task blocks. Subjective workload ratings were collected following each task block. Performance measures included accuracy and response time.

Thermal Readings

A MikronScan 7200V Thermal Camera (Mikron Infrared, Inc., Oakland, NJ) was used to measure changes in facial temperature. The MikronScan 7200V is a non-contact, high sensitivity infrared radiometer. The range of measurable temperature of the camera is 0°C to 500°C (32.0°F to932.0°F) with the sensitivity/NETD of 0.08°C.

Several thermal readings were collected, a baseline measurement and block thermal readings. A baseline assessment was collected following a 15 minute stabilization period to allow participants to acclimate to the room. Baseline images were collected for 5 seconds with the middle three seconds of data used to determine baseline readings to eliminate camera start-up and shut down effects. Thermal data was collected continuously for the duration of each task block, again excluding the first and last 10 seconds of data to eliminate camera start-up and shut down effects. For all images, the camera was positioned 45 cm in front of the participant allowing for full visualization of the face and neck, and images were sampled at 1Hz.

Analysis of the images was conducted using Regions of Interest (ROIs). ROIs were superimposed over the nose and forehead (Figure 2.2). Nose ROIs did not include either nostril as air flow during breathing may affect thermal readings. Forehead ROIs included the part of the face between the eyebrows and hairline (Figure 2.2). For each frame of data, mean, minimum, and maximum temperatures were collected. Thermal readings were adjusted using the baseline image prior to statistical analysis.

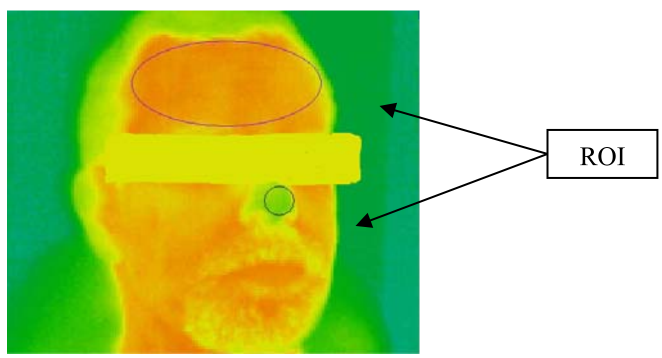


Figure 2.2

Thermal Image with ROI

Subjective Workload Assessment Tools

Two subjective workload assessment tools, the Modified Cooper Harper Scale (MCH) and the Subjective Workload Assessment Technique (SWAT), were used in this study. These tools were chosen as they are both common subjective workload measurement tools, to compare differences in a multidimensional and unidimensional scale, and to investigate which tool was more closely related with thermal readings. Further, the SWAT has a similar ability to measure subjective mental workload as NASA-TLX, though it is simpler to apply and score. Lastly, the current study was focused on the cognitive demands or attention resources demanded by a specific task, and Rubio et al. (2004) recommends using SWAT in such cases.

Modified Cooper Harper Method (MCH). The MCH scale was developed by Wierwille and Casali (1983) to measure subjective mental workload associated with cognitive and perceptual tasks (Appendix A). The MCH assesses workload in systems other than those where the human operator performs motor tasks; namely, where perceptual, mediate, and communications activity is present (Wierwille and Casali, 1983). The MCH is a unidimensional scale in which a series of questions directly leads to a single rating that ranges from 1 (very easy) to 10 (impossible). The advantages of the MCH include ease of use and reduction in administrative work.

Subjective Workload Assessment Technique (SWAT). SWAT (Reid and Nygren, 1988) is a multidimensional method that combines ratings of three different scales (Appendix B): (1) time load, reflects the amount of spare time available in planning, executing, and monitoring a task; (2) mental effort load, assesses how much conscious mental effort and planning are required to perform a task; and (3) psychological stress load, measures the amounts of risk, confusion, frustration, and anxiety associated with task performance. Researchers have found that SWAT ratings are less variable than MCH ratings (Warr, 1986), more sensitive to changes in difficulty of tracking tasks than MCH ratings (Kilmer et al, 1988), and is more sensitive to individual differences (Nygren, 1991). This study will use a modified version of the SWAT using three visual-analog scales (VAS). The VAS version of SWAT has been shown to be more sensitive in measuring moderate levels of mental workload (Luximon & Goonetilleke, 2001).

Performance measures

Accuracy rate and reaction time were recorded. Accuracy rate was defined as the percentage of correctly answered questions for each block. Reaction time was measured from the time of question presentation to the time of participant response via the keyboard.

2.6.6 PROCEDURE

Participants completed informed consent documents prior to data collection, followed by a demographic questionnaire and rested in a seated position for the 15 minute acclimation period, and the baseline thermal image recorded. The first task was block presented, and the subjective measures completed during the resting period. This process continued until all task blocks were completed. At the completion of the experiment, participants were compensated for their time.

2.6.7 DATA ANALYSIS

Appropriate descriptive statistics were calculated for each dependent variable (e.g., mean, Standard deviations). Delta facial thermographic readings were obtained subtracting the baseline reading from the recorded task block temperature for each frame. For the thermal data two representations were used. First, nose and forehead temperatures associated with the key question were analyzed, and second, rates of changes in nose and forehead temperatures analyzed. Regression model was developed to predict rates accuracy and the other dependent variables.

A repeated measures ANOVA was used to assess the effects of test block and gender on delta facial thermographic readings, delta nose temperature standard deviation, participant reaction time (RT), response accuracy, and MCH and SWAT ratings of mental workload.

Tukey's HSD post hoc tests were used where appropriate. Correlations were computed between each of the dependent variables. All findings were considered significant at an alpha of 0.05. The SAS system V9.1 for windows was used for all statistical analyses.

2.7 RESULTS

Descriptive statistics for each of the dependent variables are provided in Table 2.3. In general, subjective workload measures (MCH and SWAT) and reactions time (RT) decreased across experimental blocks, while accuracy and delta nose temperature (ΔNT) increased.

Table 2.3

Descriptive statistics for the dependent variables (values are in mean (standard deviation))

Block	ΔNT (^oC)	RT (sec)	Accuracy (%)	MCH	SWAT (mm)	ΔFT (^oC)	ΔNT S.D.
1	-1.02	2.15	90.88	2.90	149.7	0.06	0.33
	(1.41)	(0.26)	(5.81)	(0.96)	(50.81)	(0.25)	(0.20)
2	-0.22	1.87	96.14	2.70	122.9	0.04	0.23
	(1.45)	(0.24)	(3.19)	(1.03)	(54.22)	(0.29)	(0.11)
3	0.08	1.88	97.03	2.35	92.4	0.02	0.23
	(1.01)	(0.51)	(3.12)	(0.98)	(68.16)	(0.31)	(0.17)
4	0.30	1.71	97.55	2.10	83.2	0.03	0.25
	(0.81)	(0.24)	(2.53)	(0.78)	(61.67)	(0.33)	(0.18)
5	0.37	1.66	97.58	1.95	61.1	0.03	0.25
	(1.05)	(0.27)	(2.66)	(0.88)	(44.16)	(0.32)	(0.16)
6	0.32	1.66	97.34	1.75	53.8	0.02	0.25
	(1.09)	(0.32)	(3.43)	(0.78)	(42.94)	(0.34)	(0.16)
7	0.31	1.63	97.86	1.80	51.9	0.05	0.22
	(1.05)	(0.29)	(2.83)	(0.89)	(50.15)	(0.35)	(0.16)

All the dependent variables were significantly affected by block, with the exception of ΔFT readings and ΔNT standard deviations (ΔNT S.D.) (Table 2.4).No gender or block-by-gender interaction effects were found.

Table 2.4

Repeated measures ANOVA results (p-values)

Dependent Variable	Block	Gender	Block * Gender
Accuracy	< **0.0001**	0.9869	0.6259
RT	< **0.0001**	0.6985	0.5753
ΔNT	< **0.0001**	0.1397	0.9587
ΔFT	0.6277	0.2651	0.4037
SWAT	< **0.0001**	0.0809	0.8408
MCH	< **0.0001**	0.2275	0.9435
ΔNT S.D.	0.1613	0.9781	0.2435

Bolded values indicate significant findings (p-value < 0.05)

2.7.1 THERMOGRAPHY

ΔNT was found to be affected by block, while ΔFT and ΔNT S.D. were not (Table 2.4). Figure 2.3 shows the trend of ΔNT and ΔFT across the seven blocks. ΔNT showed an increasing trend while ΔFT remained stable. ΔNT in block 1 was found to be significantly lower than the other blocks (Table 2.5). Observing the ΔNT trend, task nose temperature approached baseline at block three, then remained higher for the remainder of the session (Figure 2.3).

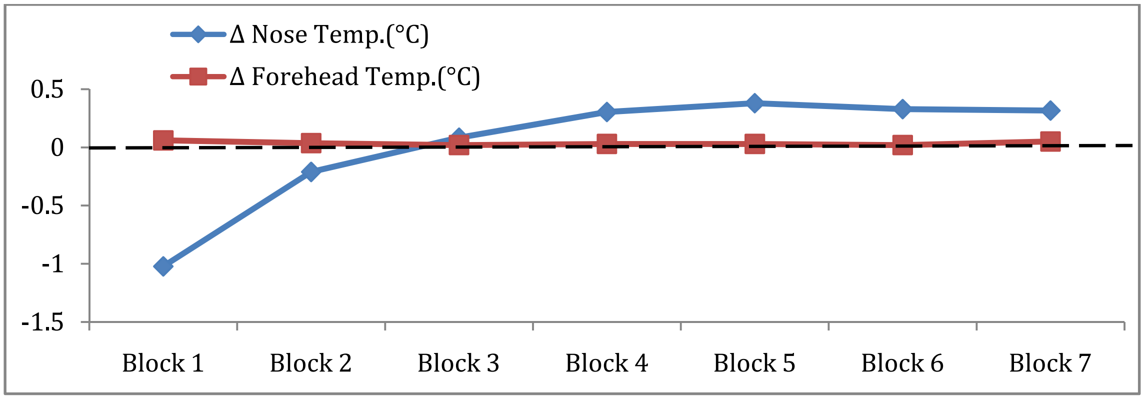


Figure 2.3

Trends of ΔNT and ΔFT

Table 2.5

Tukey's post hoc comparisons for ΔNT

Dependent Variable	Block	Mean	Group	
ΔNT	B1	-1.02 °C	A	
	B2	-0.21 °C		B
	B3	0.08 °C		B
	B4	0.30 °C		B
	B5	0.37 °C		B
	B6	0.32 °C		B
	B7	0.31 °C		B

A U-shape profile was observed for ΔNT associated with the key question presented in each question set within each task block. Further, the temperature increased in a curvilinear relationship across blocks with a remarkable ascending trend blocks 1 through 4 (Figure 2.4). The ΔNT regression slope was calculated for each subject in each block to test whether slopes were different across the task blocks by a repeated measure ANOVA. While no significant findings were found, there was a trend for differences across blocks (p-value = 0.087).

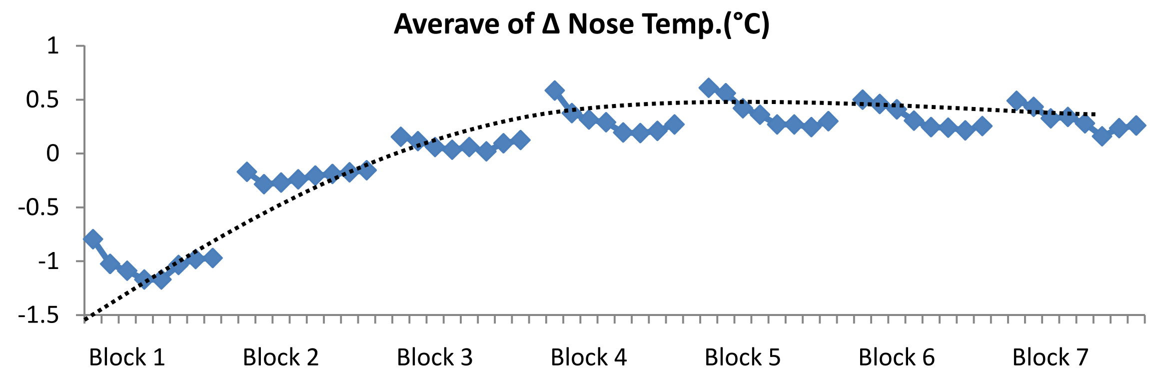


Figure 2.4

Trends for each key question in each block on average

2.7.2 SUBJECTIVE WORKLOAD MEASURES

A strong, and significant, correlation was found between the MCH and SWAT (0.99) with p-values of at least 0.05. Subjective workload assessments were significantly affected by block (Table 2.4). Subjective workload assessments were found to be similar between blocks 2 and 3 and across blocks 4 through 7. Block 1 was rated significantly higher than the other blocks for MCH ratings. MCH ratings decreased rapidly until block 4, then began to flatten for the remaining three blocks (Figure 2.5). Blocks one and two were similar for SWAT ratings. SWAT ratings followed a similar trend as MCH ratings, though the rapid decrease continued through block 5.

Table 2.6

Tukey's post hoc comparisons for subjective workload assessments

Dependent Variable	Block	Mean	Group			
SWAT	B1	149.70	A			
	B2	122.90	A	B		
	B3	92.40		B	C	
	B4	83.15			C	D
	B5	61.05			C	D
	B6	53.80				D
	B7	51.85				D
MCH	B1	2.90	A			
	B2	2.70		B		
	B3	2.35		B	C	
	B4	2.10			C	D
	B5	1.95			C	D
	B6	1.75				D
	B7	1.80				D

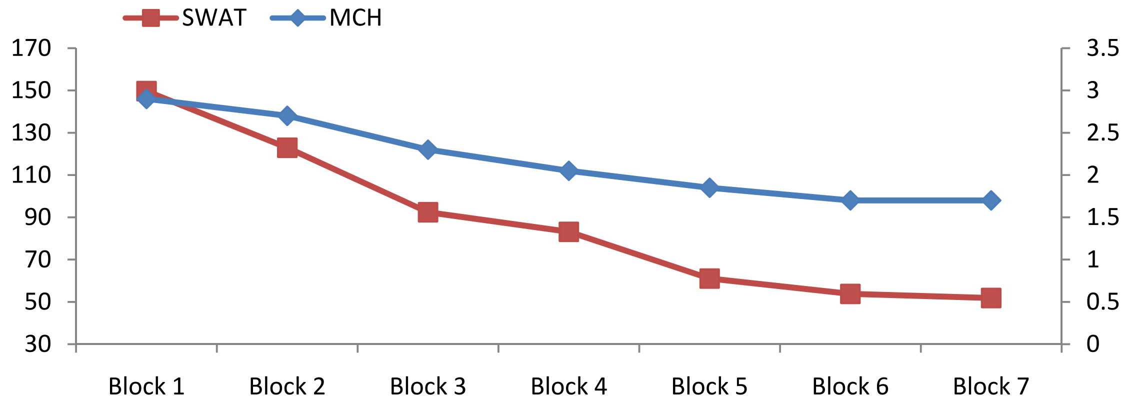


Figure 2.5

Trends of the MCH and SWAT

2.7.3 PERFORMANCE MEASURES

A strong, and significant, correlation was found between accuracy and RT (-0.92), indicating accuracy increased while RT decreased across the task blocks. Accuracy and RT were significantly affected by block (Table 2.4). For accuracy, block 1 was significantly lower than the other blocks and similar across blocks 2 through 7 (Table 2.7). Accuracy increased steeply across blocks 1 to 3 and fluctuated slightly across the remaining blocks (Figure 2.6). Block 1 was also found to be significantly lower than the other blocks, and blocks 2 and 3 and blocks 4 to 7 were found to be similar for RT. RT decreased steeply across blocks 1 to 3 and then remained similar level across blocks 4 through 7.

Table 2.7

Tukey's post hoc comparisons for subjective workload assessments

Dependent Variable	Block	Mean	Group		
Accuracy	B1	90.88 %	A		
	B2	96.14 %		B	
	B3	97.03 %		B	
	B4	97.55 %		B	
	B5	97.58 %		B	
	B6	97.34 %		B	
	B7	97.86 %		B	
RT	B1	2.15 sec.	A		
	B2	1.87 sec.		B	
	B3	1.88 sec.		B	
	B4	1.71 sec.		B	C
	B5	1.66 sec.			C
	B6	1.66 sec.			C
	B7	1.62 sec.			C

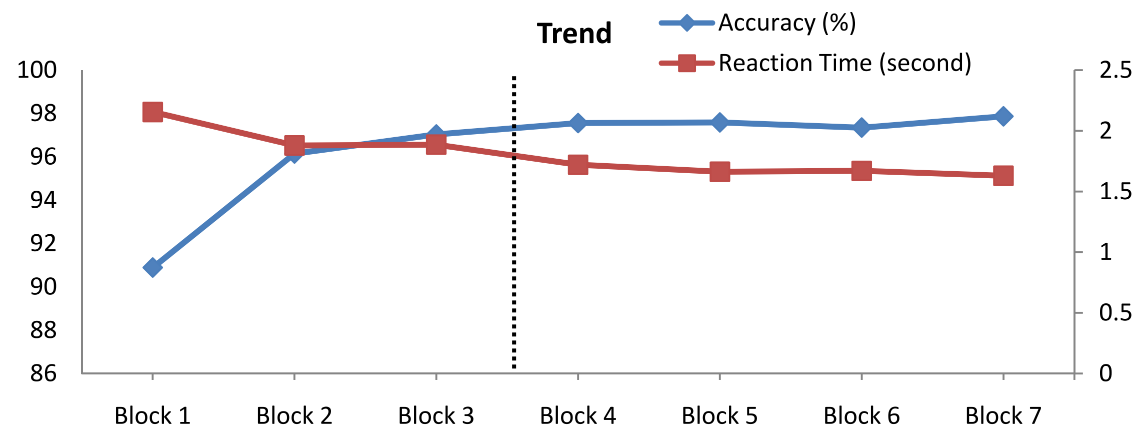


Figure 2.6

Trends of accuracy and RT

2.7.4 RELATIONSHIP AMONG MEASURES

Strong correlations, greater than ±0.80, were found between each of the dependent variables, with p-values of at least 0.05 (Table 2.8), excluding ΔNT S.D and ΔFT. Block 1 was significantly different from the other blocks for ΔNT (Table 2.5) and MCH ratings (Table. 2.6), while ΔNT and subjective workload measures were found to be similar across blocks 4 to 7 in statistical result. Figure 2.7 also shows the trend of the dependent measures across the blocks. As can be seen by this figure, subjective mental workload ratings decreased while ΔNT increased then stabilized across blocks. Figure 2.8 shows the relationship between ΔNT and performance measures. ΔNT and performance increased (accuracy) or decreased (RT) across blocks 1 through 3 and then remained similar level across blocks 4 to 7. SWAT and MCH ratings correlated strongly with ΔNT, greater than -0.90. Subjective workload measures were correlated with performance, greater than ±0.80, strongly and significantly. In general, for all trends, all measures were found to change more rapidly across blocks 1 to 3, than the remaining blocks.

Table 2.8

Correlation coefficients for all significant correlations

Variable	ΔNT	Accuracy	RT	MCH	SWAT	ΔNT S.D.
ΔNT	1.00					
Accuracy	**0.98**	1.00				
RT	**- 0.96**	**-0.92**	1.00			
MCH	**- 0.90**	**-0.80**	**0.94**	1.00		
SWAT	**- 0.93**	**-0.85**	**0.95**	**0.99**	1.00	
ΔNT S.D.	- 0.79	**-0.89**	0.72	0.54	0.63	1.00

Bolded values denote significant correlation coefficients (p-value < .05)

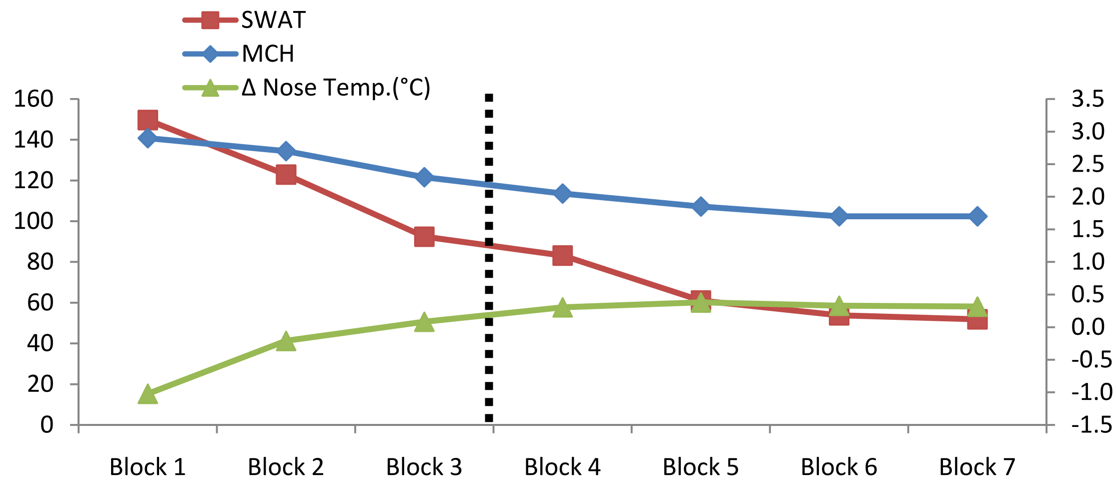


Figure 2.7

Trends of ΔNT and subjective workload measures

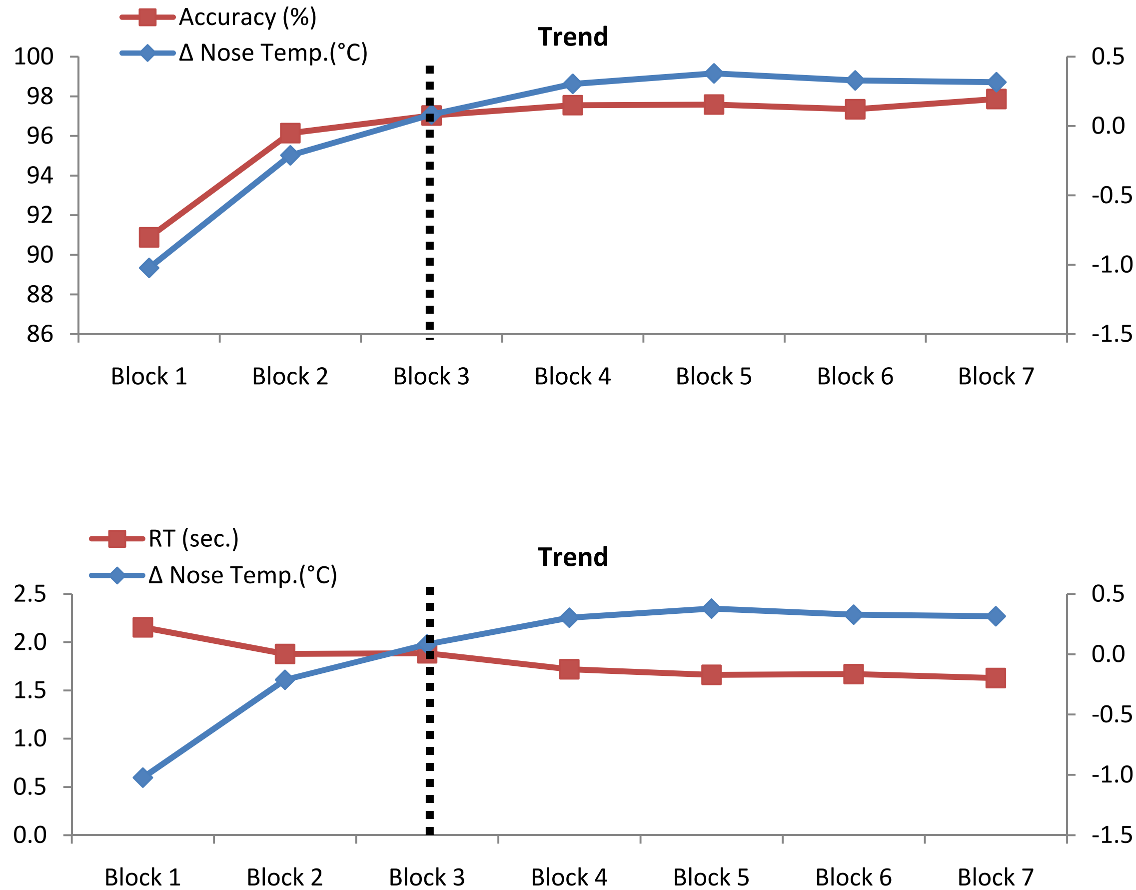


Figure 2.8

Trends of ΔNT and performance measures

2.7.5 REGRESSION MODEL

Regression was used to develop a regression model to predict accuracy using the other dependent variables (thermal variables and subjective ratings). Stepwise model building was used to develop the model with significant levels of entry and exit set to 0.05. The resultant model included only ΔNT, as provided in equation 1 below.

$$\hat{Y} = 96.2238 + 4.76709 x_i$$

Where x_i =ΔNT

$\hat{Y}$ = accuracy

Equation 2.1

Regression model

This model (Equation 2.1) indicates that accuracy increased by 4.76709 percent for each 1 degree increase in ΔNT on average, R^2 = 0.9541, indicating about 95% of the total variation in accuracy was explained by change in nose temperature (model p-value = 0.0002).Table 2.9 provides an illustration of model performance for each block.

Table 2.9

Regression model performance

Block	Obtained Value	Predicted Value
1	90.9	91.3
2	96.1	95.2
3	97	96.6
4	97.6	97.7
5	97.6	98.0
6	97.3	97.8
7	97.9	97.7

2.8 DISCUSSION

This experiment employed an alphabet arithmetic task to validate the efficacy of using thermography to assess mental workload and learning progression. This study quantified the relationship between thermal readings, participants' performance, and subjective workload assessment ratings. Six hypotheses were tested and all were supported by the results. Discussion of the findings and the relationships among the dependent variable categories, in particular, are presented below.

2.8.1 TEST HYPOTHESES

Hypothesis 1: Performance will be correlated with mental workload, specifically nose temperature readings.

Findings for this study supported this hypothesis. Thermography was strongly correlated with both performance measures captured in this study, accuracy and reaction time (RT), supporting the results found by Kang et al. (2006). Investigation of the trend in performance measures and ΔNT found that ΔNT and accuracy increased across task blocks, while RT decreased. As previous research had indicated that learner's mental workload decreases with improved performance (Jex, 1988; Tsang and Vidulich, 2006) due to learning/training, these findings provide evidence to support the utilization of thermography to quantify mental workload.

Thermography and performance measures were affected by mental workload during learning of this novel task. Performance was found to have a strong and significant correlation with subjective workload measures, indicating imposed mental workload decreased through learning progression. The first block showed significantly lower accuracy and ΔNT than the other blocks, though no differences were noted between the remaining blocks. This indicates that block 1 was where participants were undergoing significant learning. RT in the first block was also significantly longer than in the other blocks. This finding demonstrated that participants needed more time to process new information and find appropriate actions to answer the questions in block 1.

A notable finding was that participants' ΔNT fluctuated less (Figure 2.9) when participants had consistent performance throughout the test blocks. It was demonstrated by a significantly strong correlation between ΔNT S.D. and accuracy, where ΔNT S.D. decreased as accuracy increased (Figure 2.10). Participants with poorer performance and higher mental workload ratings early in the test session had a much different ΔNT profile with larger swings in their nose temperatures in general (Table 2.3). These findings provide preliminary evidence that thermography is not only sensitive to learning but also to performance differences in individuals, and that using thermography to determine training based on performance may be viable.

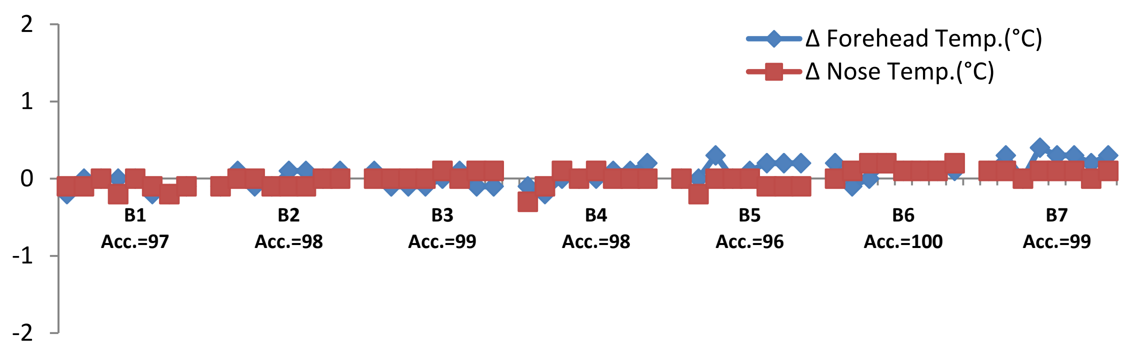


Figure 2.9

A participant's thermal readings in each block (Accuracy rate %)

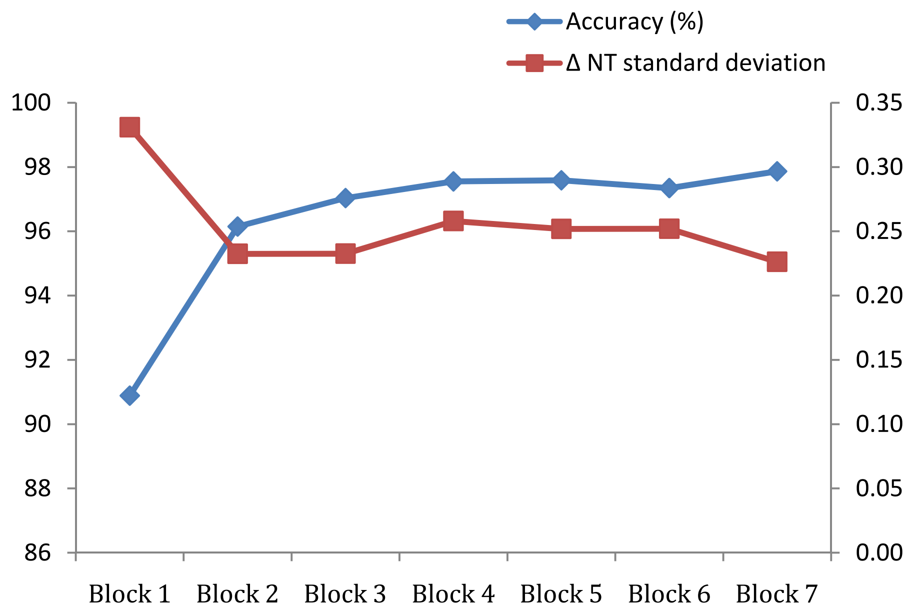


Figure 2.10

Trend of accuracy and nose temperature standard deviations across task blocks

Another remarkable finding was a U-shape profile for ΔNT that appeared within each block (figure 2.4), indicating that participants' mental workload/stress decreased until near the end of the block then began to increase again. The rest between the blocks, or even the completion of the subjective workload ratings during this rest period, resulted in an increased initial mental workload, which dissipated after a "warm-up" period. These findings were consistent across all blocks, though the absolute magnitudes of the temperatures were different.

Hypothesis 2: Nose temperature will be correlated with subjective measures of mental workload.

This hypothesis was supported by the results. Subjective workload ratings were highly, and significantly, correlated with thermal readings, supporting the hypothesis that thermography can be used to objectively quantify mental workload as a non-invasive measure. Consistent with participant perceptions, thermography indicated that significant learning occurred in block 1 (high workload) since temperature values were significantly lower in block one than the other blocks, same with previous research (Kang et al., 2006; Kang and Babski-Reeves, 2008).

Hypothesis 3: Nose temperature will be affected by mental workload changes associated with learning. However, forehead temperature will remain mostly constant.

Results for this study supported this hypothesis. ΔNT increased with improving performance, while forehead temperature remained constant across the test blocks. These findings coincided with other research reporting that nose temperature decreased when mental workload was imposed on participants (Veltman and Vos, 2005; Kang et al., 2006;Kang and Babski-Reeves, 2008), while forehead temperature was stable in the presence or absence of mental workload (Stoll 1964; Veltoman and Vos, 2005; Kang and Babski-Reeves, 2008).Nose temperature increases when mental workload is lower, such as during rest periods between cognitive tasks (Veltoman and Vos, 2005, Kataoka et al., 1998), indicating nose temperature increases with decreasing mental workload.

Hypothesis 4: Performance will have an asymptotic relationship with time (i.e., learning).

This hypothesis was supported by the results. Participants' performance was also improved over time, indicating learning acquisition (Salas et al, 2006). However, accuracy and RT showed that there was no significant improvement after block 3 because the experiment provided a simple and repetitive task so that participants achieved their best performance within a short period (Figure 2.6) (Kang and Babski-Reeves, 2008).

Hypothesis 5: Mental workload will decrease with time (i.e., learning).

This hypothesis was supported by the results. Subjective workload measures, the MCH and SWAT, found that participants' mental workload decreased while the nose temperature increased with learning acquisition. Learning or training reduces a novice's mental workload level by skill or knowledge acquisition (Moray, 1988; Jex, 1988; Wickens and Hollands, 2000, Kang et al., 2006; Kang and Babski-Reeves, 2008). The three measures in this study showed mental workload decrease due to learning, supporting the previous findings (Tsang and Vidulich, 2006; Jex, 1988).

Hypothesis 6: No gender differences will be found.

This hypothesis was supported by the results. No gender differences were found for any of the dependent variables, indicating males and females had similar mental workload and performance levels during task performance (Kang et al., 2006; Kang and Babski-Reeves, 2008).

2.8.2 SUBJECTIVE WORKLOAD ASSESSMENT COMPARISONS

A uni-dimensional mental workload assessment tool (the MCH) and a multidimensional mental workload assessment tool (SWAT) were used to determine if a single tool was more appropriate for these types of tasks, and, more importantly, to determine if one tool was more closely correlated with thermographic readings. SWAT ratings were found to have stronger correlations with thermography and performance measurements than the MCH, although these differences may be negligible since MCH ratings were also highly and significantly correlated with thermography and performance measures. Based on participant responses, however, the SWAT ratings are easier to use and thereby may be the more practical choice for tasks that require learning and decision making. The MCH instructions also complicate the rating process, which may impact ratings, though this was not necessarily found in this study.

2.8.3 SUFFICIENT TRAINING TIME

Skill mastery, or practice, occurred during blocks 2 and 3, as evidenced by thermography and performance measures. In these measures a “ceiling effect” was observed (Figure 2.8). Performance measures continued to improve across these blocks with increases in nose temperature.

Results in the study indicated that sufficient learning occurred at block 3.No further improvements in performance were observed from blocks 4 to 7. Interestingly, ΔNT increased for two blocks beyond the skill mastery blocks (blocks 2 and 3), then decreased slightly. This finding indicates that for this particular task, participants experienced overtraining and may have begun to focus on items outside of the task (multitasking) due to the task becoming easier. Another possible explanation is that workload levels may have become too low from extremely low task difficulty which has been found to be related in increased mental workload (Figure 2.11) (de Waard, 1996). de Waard suggests that high or low demands result in high workload and poor performance. Participants could have been experiencing boredom and fatigue (due to task duration). ΔNT approached baseline in block 3, again supporting that learning had occurred at the completion of this block (Figure 2.3).

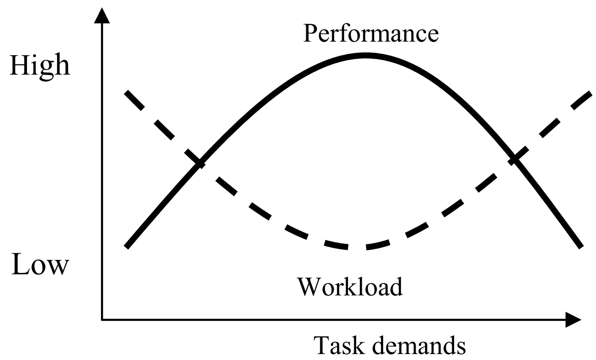


Figure 2.11

Workload, performance, and task demands relationship (adapted from de Waard, 1996)

Similar trends and findings were observed in the subjective workload measures, supporting that skill mastery occurred in blocks 2 and 3. MCH and SWAT ratings were significantly higher in block 1, then were similar across blocks 2 and 3 (the skill mastery blocks), then were similar for the remaining blocks, though different from the first 3 blocks.

Since all participants performed a novel task in this experiment, participants likely were experiencing a certain level of mental workload from outside influences (outside of the research study, such as anxiety or emotional stress) when presented with the study environment and the study task. However, the return to baseline in nose temperature and decreased mental workload indicates that the session length was sufficiently long to allow participants to acclimate to the environment and the study, as has been found previously (Gaillard, 1993; Wickens et al., 1998). These findings support the conclusion that sufficient training time for the task in this study was three blocks (about 25 minutes), one block devoted to training, and two blocks devoted to skill mastery.

2.8.4 LIMITATION OF SUBJECTIVE WORKLOAD AND PERFORMANCE MEASURES

Subjective mental workload measurements were found to correlate strongly with performance measurements (greater than ± 0.8), supporting previous findings indicating that performance and mental workload levels are related (Tsang and Vidulich, 2006; Kang et al., 2006). Similar findings for subjective workload ratings were found for block 1 as for performance and nose temperature. Workload levels during block one were significantly higher than the other blocks indicating that learning was occurring. However, subjective workload measures sustained a decreasing trend across the task blocks, which is different from the trends for performance and thermography. The findings indicate that participants' subjective workload ratings did not coincide with "true" mental workload and the participants' perceptions of mental workload differed from their performance (Tsang and Vidulich, 2006; Kang et al., 2006; Jex, 1988; Andre and Wickens, 1995). Since the task was simple and repeated, it is likely that participants may have perceived decreasing task difficulty without detecting the boredom or fatigue that may have been resulting, consistent with previous findings (Gaillard, 1993; Wickens et al., 1998; de Waard, 1996). The experiment duration was not sufficient to identify the long term trend in performance was affected by other factors.

Thermogarphy showed that there were a learning stage and skill mastery stage. It also demonstrated workload increased when participants had overtraining by decreasing trend across blocks 5 through 7(Figure 2.3). Performance and subjective mental workload measures confirmed that there was a learning, skill mastery, and task performance stage. However, performance measures and subjective ratings were not able to identify these stages in real-time. For instance, there is a need to spend limited time on training to minimize costs. Thermography, on the other hand, would be an alternative measures where trends in nose temperature can be monitored to identify when sufficient training has occurred.

Thermography can record facial temperature readings continuously, performance and subjective measures are difficult, if not impossible, to collect and track in real time. Thermography can illustrate mental workload fluctuations during training (Figure 2.4), allowing trainers to decide when sufficient training has occurred on an individual basis. Again, subjective or performance measures are limited in their ability to do this.

2.9 FUTURE WORK AND LIMITATIONS

While this study supports the use of thermography for mental workload assessment for decision making tasks, other tasks need to be investigated to illustrate the robustness of this tool for assessing mental workload. Also, various environmental conditions should be studied to quantify the effects of directly blowing air from windows or air vents on the findings. Also, thermography needs to be investigated under a longer experiment design to test an overtraining effect on participants for all dependant variables.

This study developed a regression model using all dependent variables and the step wise model building technique. The model provides support for using thermography based on performance, as 95% of the variability in accuracy was explained by changes in nose temperature. Further study is needed to build an accurate regression model including individual differences.

The condition of the participants needs to be controlled during experiment. For example, some participants conducted the task right after their classes, which could affect participants' initial mental workload and physical fatigue for their learning progression. Further, research needs to investigate the participants' physical and mental fatigue during the cognitive task because participants did not understand that mental workload is affected by fatigue.

This study provided 4 seconds for participants' to respond to the question presented to them. Further study needs to investigate an effect between different response times (2 or 10 seconds), as this may have also affected mental workload levels.

Research needs to investigate the utilization of thermography with a complex task that provides more complicated mental workload and circumstance on learners.

2.10 CONCLUSION

This study verified thermography is a valid, objective, and non-invasive mental workload measure. Performance and subjective mental workload findings support the use of thermography to quantify mental workload for simple novel tasks. This study demonstrated the utility of using thermography to successfully identify sufficient training time by monitoring fluctuations in nose thermal temperatures during learning. This indicates that thermography is more sensitive to slight changes in mental workload than other, more traditional mental workload measures. Thermography is a more practical observation method as it does not interfere with task performance or require the attachment of equipment to the human. Costs associated with using and analyzing thermal data need to be considered, however.

2.11 REFERENCES

Andre, A.D., & Wickens, C.D. (1995). When users want what's not best for them. *Ergonomics in Design:The Quarterly of Human Factors Applications 3(4)*, Oct. 10-14(5).

Alvager, T., Anderson, E., French, V.A., Putman, G., & Shi, L. (1999). An artificial neural network study of the relationship between arousal, task difficulty and learning, *IJCNN '99. International Joint Conference on Neural Networks*, 5. 3618 – 3621.

Boucsein, W. (1993). Psychophysiology in the work place-Goals and methods. In P. Ullsberger (Ed.) *Psychophysiology of mental workload* (pp. 35-41). Berlin: Bundesanstalt fur Arbeitmedizin.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games, *Applied Artificial intelligence*, 16, 555-575.

De Waard, D. (1996). *The measurement of driver's mental workload*. PhD Dissertation,University of Groningen, Netherlands.

Drummond, P.D., Camacho, L., Formentin, N., Heffernan, T.D., Williams, F., & Xekas, T.E. (2003). The impact of verbal feedback about blushing on social discomfort and facial blood flow during embarrassing tasks.*Behaviour Research and Therapy,* 41, 413-425.

Gaillard, A.W.K. (1993). Comparing the concepts of mental load and stress. *Ergonomics*, 36, 991-1005.

Gawron.V.J. (2000).*Human performance measures handbook.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Genno, H.[a], Saijo, A., Yoshida, H., Suzuki, R., & Osumi, M. (1997). Non-contact method for measuring facial skin temperature. *International Journal of Industrial Ergonomics, 19,* 147-159.

Genno, H.[b], Ishikawa, K., Kanbara, O., Kikumoto, M., Fujiwara, Y., Suzuki, R., et al. (1997). Using facial skin temperature to objectively evaluate sensations. *International Journal of Industrial Ergonomics, 19,* 161-171.

Gevins, A., & Smith, M.E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, 4, 113-131.

Green, J., & Shellenberger, R. (1991). *The dynamics of health and wellness: A Biopsychosocial Approach*. FortWorth: Holt, Rinehart and Winston, Inc.

Guyton, A.C., & Hall, J.E. (2006). Text book of medical physiology (11[th]). Philadelphia, PA: Elsevier Inc.

Hankins, T.C., & Wilson, G.F. (1998). A comparison of heart rat, eye activity, EEG, and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environmental medicine*, 69, 360-367.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Handcock & N. Meshkati, (Eds.),*Human Mental Workload*(pp. 139-178).Amsterdam: North-Holland.

Hockey, G.R.J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. In K.R. Boff, L. Kaufman & J.P. Thomas (Eds.), *Handbook of perception and human performance*, Vol. II, 44-1- 44-49.

Jex, H. R. (1988). Measuring mental workload: Problems, progress, and promises, In P.A. Hancock & N.Meshkati, (Eds.), *Human Mental Workload* (pp. 5-39). New York: North-Holland.

Johnson, D.F., & Haygood, R.C. (1984). The use of secondary tasks in adaptive training. *Human Factors*, 26(1), 105-108.

Jewell, E.J., & Abate, F. (Eds.). (2001). *The new Oxford American dictionary (1st ed.)*. New York: Oxford University Press.

Kaber, D. B., Riley, J. M., Tan K., & Endsley, M. R.. 2001, On the Design of Adaptive Automation for Complex Systems.*International Journal of Cognitive Ergonomics*, 5, 37-57.

Kahneman, D. (1973). *Attention and effect*. Englewood Cliffs, NJ: Prentice-Hall.

Kang, J. and Babski-Reeves, K. (2008). Detecting mental workload fluctuation during learning a novel task by thermography. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual meeting* , September 22-26, 2008, New York, NY.

Kang, J., McGinley, J., McFadyen, G., & Babski-Reeves, K. (2006). Determining learning level and effective training times using thermograhy. *Proceedings of 25thArmy Science Conference*, Orlando, FL.

Kataoka, H., Kano, H., Yoshida, H., Saijo, A., Yasuda, M.,& Osumi, M. (1998). Development of a skin temperature measuring system for non-contact stress evaluation. *Proceedings of the 20th Annual International conference of the IEEE engineering in Medicine and Biology Society*, 20 (2), 940-943.

Kilmer, K.J., Knapp, R., Borresen, R., Bateman, R.P., & Malzahn, D. (1988). A comparison of the SWAT and modified Cooper-Harper scales. *Proceedings of the Human Factors 32nd Annual Meeting*. 155-159.

Kramer, A.F. (1987). Special issue on cognitive psychophysiology. *Human Factors*, 9, 129-127.

Landauer, T. (1995). *The trouble with computers*. Cambridge, MA: MIT Press.

Logan, G.D., & Klapp, S.T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity?. *Journal of Experimental Psychology: Learning Memory, and Cognition*, 17, 179-195.

Luximon, A., & Goonetilleke, R. S. (2001). Simplified subjective workload assessment technique. *Ergonomics, 44*, 229-243.

Montgomery, D.C., & Runger, G.C. (2003). Applied statistics and probability for engineers (3rd ed.). USA. John Willey& Sons.

Moray, N. (1988). Mental workload since 1979. In D.J. Oborne (Ed.), *International Reviews of Ergonomics: Vol. 2. Current trends in human factors research and pratice (International reviews of ergonomics)* (pp. 123-150). London: Taylor & Francis.

Moray, N. (1979).*Mental Workload: Its Theory and Measurement*. New York: Plenum.

Montgomery, D.C., & Runger, G.C. (2003). Applied statistics and probability for engineers (3rd ed.). USA. John Willey& Sons.

Naemura, A., Tsuda, K., & Suzuki, N. (1993). Effects of loud noise on nasal skin temperature. *The Japanese Journal of Psychology*, 62, 51-54.

Nygren, T.E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors*, 33, 17-33.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. BoŠ , L. Kaufman, & J. Thomas (Eds.), *Handbook of Perception and Human Performance : Volume II. Cognitive Processes and Performance* (pp. 42.1-42.9).New York: Wiley.

Pavlidis, J. L., Levine, J., &Baukol, P. (2000). Thermal Imaging for Anxiety Detection, *Proceedings of the 2000 IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, Hilton Head Island, South Carolina, 104-109.

Reid, G.B., & Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock & N. Meshkati. (Eds). *Human Mental workload* (pp. 185-213), Amsterdam: North Holland.

Rasmussen, J. (1983). Skills, rules, knowledge: Signals, signs, and symbols and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13, 257-267.

Rasmussen, J (1986). *Information processing and Human-machine interaction: An approach to cognitive engineering*. New York: Elsevier.

Rasmussen, J (1993). Deciding and doing: Decision making in natural contexts, In G.Klein, J, Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 158-171). Norwood, NJ: Ablex.

Rubio, S., Díaz, E., Martín, J, &Puente, J.M. (2004).Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods.*Applied Psychology: An international review,* 53, 61-86.

Salas, E., Wilson, K.A., Priest, H.A., & Guthrie, J.W. (2006). Design, delivery, and evaluation of training systems, In G. Salvendy (Ed.). *Handbook of Human Factors and Ergonomics* (pp. 234-268). New Jersey: John Wiley& Sons.

Sarter, N., & Woods, D.D. (1997). Teamplay with a powerful and independent agend: A corpus of operational experiences and automation surprises on the Airbus A-320. *Human Factors*, 39, 553-569.

Sanders, M.S., & McCormick E., J. (1993).*Human factors in engineering and design (7th ed.).* New York, McGRAW-Hill, Inc..

Sheridan, T. (1980). Mental workload: what is it? Why bother with it?. *Human Factors Society Bulletin*, 23, 1-2.

Shute, V.J., Gawlick, L.A., & Gluck, K.A. (1998). Effects of practice and learner control on short- and ling-term gain and efficiency. *Human Factors*, 40, 296-310.

Slocum, G.K., Williges, B.H., & Roscoe, S.N. (1971). Meaningful shape coding for aircraft switch knobs. *Aviation Research Monographs,* 1, 27-40.

Stoll, A.M. (1964). Techniques and uses of skin temperature measurements. *Annals of The New York Academy of Sciences: Thermography and Its Clinical Applications*, 121(1), 49-56.

Tasng, P.S., & Vidulich M.A. (2006). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics (3rd ed*.) (pp. 243-268). Hoboken, NJ: John Wiley& Sons.

Thompson, L.A., Williams, k.l., L'Esperance, P.R., & Cornelius, J. (2001). Context-dependent memory under stressful conditions: The case of skydiving. *Human Factors*, 43, 611-619.

Trujillo, A.C. (1998). Pilot mental workload with predictive system status information. *Proceedings of the 4th Symposium on Human Interaction with Complex Systems*, Fairborn, OH,73-80.

Veltman, H.J.A., & Vos, W.K. (2005). A methodology for non-intrusive mental workload measurement in simulated and on-road driving. *Proceedings of the 11th International conference on Human-computer interaction*, Las Vegas, NV.

Vernet-Maury, E., Deschaumes-Molinaro, C., Delhomme, G., & Dittmar, A. (1993). Autonomic nervous system activity and mental workload. In P. Ullsberger (Ed.) *Psychophysiology of mental workload* (pp. 42-48). Berlin: Bundesanstalt fur Arbeitmedizin.

Warr, D.T. (1986). A comparative evaluation of two subjective workload university measures: the subjective assessment technique and the modified Cooper-Harper Rating. *Masters Thesis*, Dayton, OH: Wright State University.

Wickens, C.D. (1979). Measures of workload, stress and secondary tasks. In N. Moray (Ed.), *Mental Workload: Its Theory and Measurement*(pp. 79-99). New York: Plenum.

Wickens, C.D., Gordon, S.E., & Liu, Y. (1998). *An introduction to human factors engineering.* New York, Addison: Wesley Longman, Inc.

Wickens, C.D., Mavor, M., Parasuraman, R., & McGee, J. (1998). *The future of air traffic control.* Washington, DC: National Academy of Science.

Wickens, C.D., & Hollands, J.G. (2000). *Engineering Psychology and Human Performance (3rd Ed.).* Upper Saddle River, NJ, Prentice_Hall Inc.

Wierwille, W.W., & Casali, J.G. (1983). A validated rating scale for global mental workload measurement applications. *Proceedings of the Human Factors Society 27th Annual Meeting*, Norfork, VA, 129-133.

Wierwille, W.W. and Connor, S.A. (1983). Evaluation of 20 workload measures using a psychomotor task in a moving-bas aircraft simulator. *Human Factors*, 25, 1-16.

Wilson, G.F. (1992). Applied use of cardiac and respiration measures: practical consideration and precautions. *Biological Psychology*, 34, 163-178.

Xie, B., & Salvendy, G. (2000). Review and reappraisal of modeling and predicting mental workload in single- and multi-task environments. *Work & Stress*, 14(1), 74-99.

Yerkes, R.M., & Dodson, J.D. (1908). The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.

Young, M.S. & Stanton, N.A. (2001). Mental workload: theory, measurement, and application. In W. Karwowski (Ed.), *International Encyclopedia of Ergonomics and Human Factors* (pp. 507-509). New York: Taylor and Francis.

CHAPTER 3

QUANTIFICATION OF OPTIMAL MENTAL WORKLOAD AND HUMAN PERFORMANCE LEVELS BASED ON THERMAL IMAGING OF THE FACE

3.1 ABSTRACT

Objective: The objective of this study is to investigate the utility of using psychophysical methods to determine workload levels that maximize performance on a cognitive task by subjective (SWAT and MCH ratings) and objective (thermography and performance) workload measures.

Methods: Twenty eight participants, 14 male and 14 female, completed 7 blocks of an alpha-numeric task. The 7 blocks consisted of adjustment session (3 blocks) and work session (4 blocks). The adjustment session was determined by psychophysical methods. Changes in nose and forehead temperature, task accuracy, reaction time, and two subjective mental workload ratings (MCH ratings and SWAT) were collected.

Results: Thermography (ΔNT) and accuracy showed that mental workload level in block 1 was higher than the other blocks while subjective workload measures showed no difference between blocks in the adjustment session, while there were no differences in the work session. High performers had a higher accuracy rate and thermal readings

(i.e.lower mental workload) than low performers. Participants identified their self-selected work pace before the completion of block 2. However, both high and low performers failed to find true optimal workload level for long term work.

Conclusion: Thermography can demonstrate whether participants find their optimal workload level for a short term cognitive task. Thermograpy provides more practical results than subjective workload measurements for determining an optimal workload level.

3.2 INTRODUCTION

Human performance is defined as “the accomplishment of a task by a human operator” (Gawron, 2000, p13). Designers and operators have realized that human performance is a significant factor in productivity and in the work of human safety. Improved productivity cannot be achieved without consideration of the operators’ safety and satisfaction because the operator is the main character during task performance. Consideration of a task’s demands on operator’s mental and physical capabilities and capacities is also critical for the improvement of system safety and productivity (Wickens and Hollands, 2000).

Since the role of mental activity; such as information processing, monitoring, controlling, and decision-making (Moray, 1988); has been increasing significantly in many tasks, superior human performance on cognitive tasks is essential to achieve a successful result (Hollnagel, 2006). Required parallel focus on multiple task-related issues, such as cognitive control function allocations and awareness of task responsibilities, may result in cognitive workload exceeding recommended levels (Kaber and Riley, 1999; Scerbo, 1996). However, methods for determining optimal human performance levels have not been thoroughly researched with respect to operator mental workload. This study investigates the relationship between mental workload and human performance and introduces a methodology to determine an optimal task performance level based on operator mental workload level.

3.3 BACKGROUND

3.3.1 PERFORMANCE AND MENTAL WORKLOAD

Mental workload work is becoming increasingly important in modern systems as operators assume roles that require the performance of complicated cognitive tasks, such as system management, decision making, and multi-system monitoring (Colle et al., 1988). Moray (1988) indicated that human performance and mental workload level have a strong relationship because it appears that human performance is improved (e.g., reduced human error, improved safety and systems, and increased productivity) when the operator reaches an optimal mental workload level.

Task performance can be divided among Rasmussen's SRK (skill, rule, knowledge) model (Rasmussen, 1983, 1986, 1993) of cognitive control (Figure 3.1). The level of performance a person has achieved depends on the experience of the worker (particularly their experience with the current situation) and the nature of the task. When exposed to a novel situation, persons do not possess rules for determining appropriate actions; therefore, they are required to perform at the knowledge-based level (analytical processing using conceptual information). Performance errors typically are associated with a lack of experience and "guessing" about appropriate actions/decisions. Rule-based performance exists when persons have been exposed to the task and are learning the task, though insufficient learning/training time has elapsed. Errors result from the failure to grasp or misinterpret the situation resulting in the application of an incorrect rule. Skill-based performance levels are achieved through extensive practice and training. In this level, people process information automatically, without interpreting and integrating the cues or thinking of possible actions. Performance errors at this level can result from inattention, missed cues, etc. resulting from expectations that are not realized. When learners reach the skill-based level, they use fewer cognitive resources and have a lower workload level (Tsang and Vidulich, 2006).

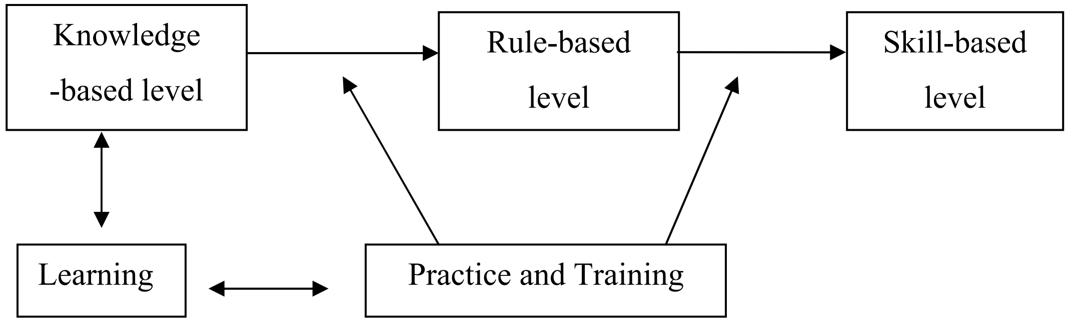


Figure 3.1

Rasmussen's taxonomy and Wickens et al task performance hierarchy

As a person moves from the knowledge-based level to the skill based level, cognitive arousal is reduced along with feelings of anxiety and stress (Conati, 2002; Alvager et al, 1999; Thompson et al, 2001), as familiarity with the task and the strategies for successful task performance are learned and retained.

Researchers have found mental workload to be a significant factor in determining human performance in complex systems (Gopher and Donchin, 1986; Colle et al, 1988; Hancock and Meshkati, 1988; Moray, 1979; O'Donnell and Eggemeier, 1986; Jex, 1988). Both low and high mental workload levels can degrade performance (Lysaght et al., 1989) and there is an optimal level of mental workload that corresponds to "optimal" performance levels (e.g. Young and Stanton, 2001; Kaber et al, 2001). Yerkes-Dodson's law (1908) presents a parabolic function that concurs with the relationship between mental workload and performance. It also demonstrates there is an optimal mental workload level (Figure 3.2) that can reduce human errors, improve system safety, increase productivity, and improve operator satisfaction (Moray, 1988).

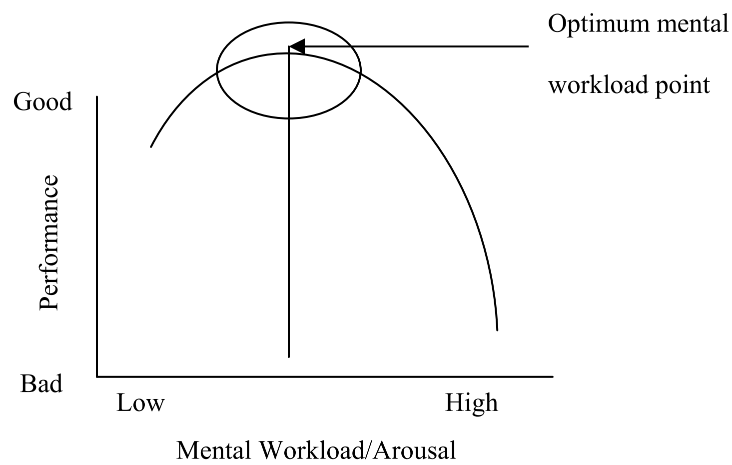


Figure 3.2

Relationship between performance and mental workload

Meister (1976) proposed a different model of the human performance-mental workload relationship (Figure 3.3). In this model, three performance regions are identified: region A is the region of low workload and high performance; region B presents a declining performance due to increased task demands; and region C is the region of a minimum performance level as a result of catastrophic high workload attributed to high task demands. In the optimal performance region (Region A), operators can easily manage task demands, and performance remains stable, without increased operator effort, as demand increases. Regardless of the relationship that individual researchers prescribe to, performance is impacted by workload level.

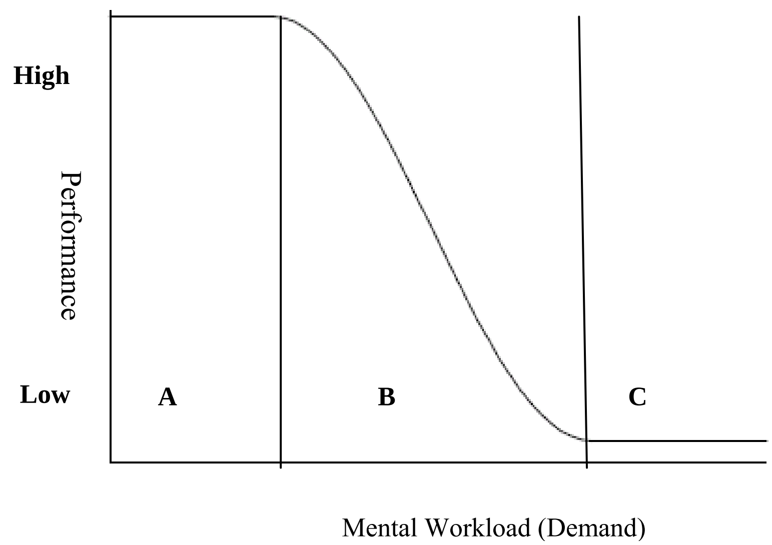


Figure 3.3

Relationship mental workload and performance based on Meister's model (1976)

High workload can demolish the quality and quantity of an operator's task performance because high workload causes operators to struggle in an attempt to maintain dynamic system states (Kaber et al., 2001). High-demand tasks may result in a mismatch between workload levels and human capacities to effectively deal with these levels (Hancock and Chignell, 1988). These assertions about an optimal workload level support the idea that an optimal performance for a specific job can be detected by measuring mental workload.

3.3.2 MENTAL WORKLOAD AND TIME CONSTRAINTS

The importance of considering time constraints during information processing is accepted by many researchers (Reid and Nygren, 1988; Hendy et al., 1997; Hancock and Chignell, 1988). For example, Zakay (1993) described the time constraint of information processing as mental workload, because operators must share their limited mental resource between processing time estimation (e.g. "how much time do I have to finish?") and the task at hand. Hancock and Chignell (1988) suggest that "mental workload should be described in terms of time, distance to desired goal, and effort". The presence of time constraints imposes a load stress on the operator, further increasing mental workload levels (Chiles and Alluisi, 1979).

The relationship between performance and time constraints is similar to that between performance and mental workload (Figure 3.4) (Johannsen, 1979). Under low workload or time constraint conditions, operators may experience boredom and fatigue due to lack of cognitive interest in control tasks. Cognitive overload may occur when operators must perform complex tasks or a large number of divergent or diverse tasks simultaneously or over a short period of time.

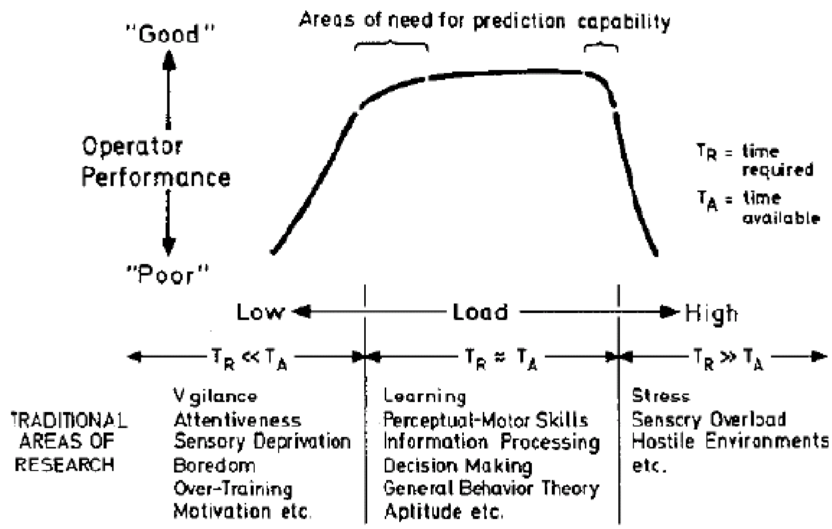


Figure 3.4

Relationship between operator performance and time constraints (Johannsen, 1979).

The most accepted relationship between accuracy and time stress is the speed accuracy trade-off (Conrad, 1956). As with mental workload and performance, there is an optimal work speed (or time constraint level) that will maximize accuracy or performance (Figure 3.5) (Wickens and Hollands, 2000). When operators are working under extreme time constraints, the probability of errors increases, but more importantly, no information will be retained by or transmitted to the operator. When too much emphasis is placed on accuracy rates, task completion times will be prolonged with little gain in accuracy and value to the task.

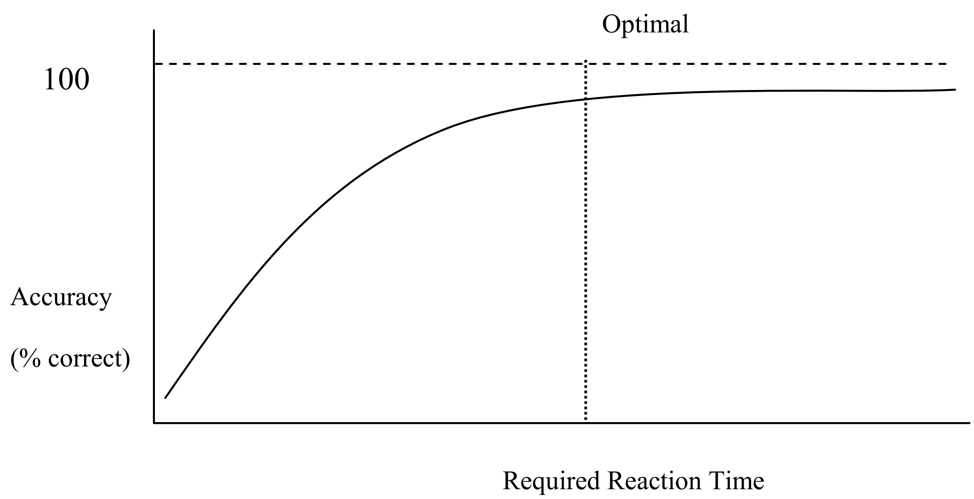


Figure 3.5

Speed accuracy trade-off

Time constraints have been found to be a major contributor to mental workload levels (Reid and Nygren, 1988). Among the various tools for measuring mental workload, subjective workload tools allow for the consideration of time stress during workload assessment most readily. Examples of these tools include the Subjective Workload Assessment Technique (SWAT) (Reid and Nygren, 1988) and the NASA Task Load Index (NASA-TLX) (Hart and Staveland, 1988).

Incorporating consideration of time stress into objective workload measures presents a challenge, as the activities that introduce workload are typically difficult to detect (such as planning and rehearsal). Wickens and Hollands (2000) observed that the covert nature of cognitive activities during task performance makes estimating the time duration of these tasks and observing their execution difficult, which complicates mental workload estimation. However, objective tools that measure responses directly related to changes in time constraints (such as blood flow) may prove to be valid and reliable mental workload assessment tools.

3.3.3 MENTAL WORKLOAD AND THERMOGRAPHY

Several measures are available for measuring mental workload and are typically classified as primary, secondary, subjective, or physiological measures. Primary and secondary task measures assume that as workload increases the additional processing requirements will degrade performance (Gawron, 2000). Literature indicates that primary performance measures have sensitivity problems in under and over load regions. Primary task measures cannot detect workload changes with increasing work demands in the under load region due to operator spare processing capacity (floor effect), with a similar ceiling effect in the over load region (O'Donnel and Eggemeier, 1986). Although secondary performance measures are more sensitive than primary performance measures, secondary measures depend on how participants interpret and determine the primary task. O'Donnel and Eggemeier (1986) also report that all performance measures have other potential problems, such confounding effects of individual differences in information-processing strategy, training, or experience which can vary from task to task. This makes choosing a particular task for assessment difficult because operators perform differently on each unique task. Workload is a subjective perception (Gailliard, 1993), therefore subjective techniques that require persons to quantify their experience of workload are favored (Tsang and Vidulich, 2006). Subjective measures, however, may not be representative of the “true” workload level because they are based only an overall work for the entire task and are vulnerable to individual differences (Johannsen, 1979). Physiological measures provide continuous and objective information about the state of an operator (Veltman and Vos, 2005), but they often require equipment to be attached to the operators, confounding the results. Therefore, alternative workload measures, which do not adversely affect the results, are needed.

By its definition, mental workload levels results in a predictable and measurable physiological reaction. Heart rate, heart rate variability, respiration rate, and other physiological responses have been used to measure mental workload levels. However, other physiological responses, particularly those associated with the autonomic nervous system (ANS) may be useful indicators of mental workload. Kataoka et al. (1998) indicate that skin temperature is a useful objective measure for analyzing mental workload because of its dependence on the ANS, which “reflects the course of information processing in the brain”. Researchers have demonstrated a high correlation between stress nose skin temperatures (Naemura et al, 1993).

Recently, Veltman and Vos (2005) and Kang et al (2006) demonstrated a relationship between mental workload and a change in facial temperature. Kang et al (2006) also demonstrated a relationship between facial thermal readings and subjective workload and performance measures. Facial temperature can be measured objectively with use of thermography, a non-invasive autonomic measure based on blood flow due to physiological reactions on the autonomic nervous system (ANS). However, studies that support or validate the use of thermography as a mental workload measure are limited. A primary benefit to using thermography to quantify mental workload is that it is a non-contact physiological assessment tool, and therefore, does not suffer from the same limitations as other physiological measures.

3.3.4 PSYCHOPHYSICAL ESTIMATION OF WORKLOADS/TASK DEMANDS

Despite our knowledge that time pressure and workload levels impact performance and there is a known “optimal” level of workload/time pressure that will maximize performance, few well documented techniques for quantifying optimal work levels or task demands exist in the literature. Psychophysics may provide a technique to determine optimal workload levels.

Psychophysics studies the relationship between physical intensities and individual perceptions of those intensities (Borg 1990). A primary assumption in psychophysical methods is that people have the ability to be ‘self-limiting’ or ‘self-protecting’. That is, when exposed to a work task, people can effectively identify maximum acceptable limits for various work task parameters (such a weights, frequencies, durations, etc.). Participants self adjust task parameters over a predetermined time (e.g., 30 minutes) until they perceive that they can perform the task over a prolonged period of time (e.g., work shift). A general adjustment time in psychophysical studies is 20 minutes and has been supported by various researchers, though both longer and shorter time periods are common (Snook, 1978; Legg and Myles, 1981; Mital, 1983; Karwowski and Yates, 1984).

Psychophysical methods may have broader utility in studies involving human work as individuals define what they are capable of performing (Grant et al., 1994). Psychophysical data is easy to interpret, making it desirable (Grant et al., 1994). Additionally, they allow for the identification of ranges of acceptable limits of workers performing tasks, which may be more effective in helping to establish safe work guidelines (Putz-Anderson and Grant, 1995).Limitations in psychophysical methods exist: training is required for reliable psychophysical estimation; errors are common with extreme conditions (Karwowksi and Yates, 1986); participants must cooperate; there is the probability that participants as operator may overestimate their capability (Ayoub and Dempsey, 1999).

Psychophysical methods have been used to determine maximum acceptable limits for a number of different types of tasks (lifting, wrist flexion/extension, etc.). While no studies were found for acceptable limits associated with cognitive tasks, the same principal may be applicable (people may be effective at identifying the optimal workload levels). This study investigates the utility of using psychophysical methods to identify acceptable mental workload levels.

3.4 OBJECTIVE

The objective of this study was to assess the utility of using psychophysical methods to determine workload levels that maximize performance on a cognitive task. Participants were allowed to adjust the pace of questions over 20 minutes of a test session, with participants performing the task at the identified work pace for the remaining time of the session (approximately 28 minutes). Thermal readings during "optimal" levels (high performers) were compared to thermal readings during "non-optimal" levels (low performers).

3.5 HYPOTHESES

Specific hypotheses investigated include:

1. Performance will be affected by block (i.e., performance will improve until reaching the optimal workload level).
2. Facial thermal readings will stabilize as the work level is adjusted toward optimal.
3. Thermal readings, performance, and subjective workload measures will be correlated.
4. Participants will be able to identify an optimal question pacing to minimize workload.
5. Gender will not affect any dependent variable.

3.6 METHODOLOGY

3.6.1 EXPERIMENTAL DESIGN

A mixed subject design was used to test for differences between optimal and non-optimal workload levels (a within subjects factor) based on starting pace (a between subjects factor) on mental workload and performance assessment measures. Mental workload was measured using thermal imaging of the face and subjective workload assessment ratings. Performance measures included reaction time and accuracy on the selected task. Many of the procedures from study one were replicated in this study. The same alpha-numeric task was used, which consisted of 7 test blocks. Participants were allowed to adjust the time between questions to identify their optimal work level. Participants were allowed to adjust the pace of the questions during the first three test blocks, equating to approximately 20 minutes. For the remaining four test blocks, participants performed the task at the selected pace identified at the end of the third test block. Exposure to starting pace (fast or slow) was counterbalanced to minimize order effects.

3.6.2 PARTICIPANTS

A total of 28 participants (14 males and 14 females) completed the study protocols. Participants ranged in age from 18 to 27 years. Participants were asked to be familiar with the study task, therefore, participants completing study 1 were recruited for this study. For those participants that did not had completed study one, a familiarization session (a complete trial as described in study 1) was completed prior to the sessions described below. Participants were randomly assigned to one of two pace groupings (based on beginning pace—see independent variables below), and each group included 14 participants (7 males and 7 females).

3.6.3 TASK DESCRIPTION

The same task as described in study 1 was used here. An alphabet arithmetic task (Logan and Klapp, 1991) was used to assess learning a novel task. In the alphabet arithmetic task, participants verified equations of the form C+2=E, D+3=G, and E+4=I. Participants were asked to perform the task by counting through the alphabet one letter at a time until the number of counts equals the digit added. Participants were asked to compute problems by adding the letters C, D, or E with a numbers 1 through 4. As before, each trial consisted of 7 test blocks containing 8 question sets, and each question set contained 12 questions. The same questions were used in each block, though their order of presentation was randomized. A single question remained in the same location throughout the testing session to allow for a consistent comparative question. A single trial was 6 minutes and 24 seconds in length and 3 minute rest periods were provided between blocks.

Participants adjusted the pacing of time between questions within each question block for the first three blocks. Question pacing was set 0.5 seconds (short pacing) and 4 seconds (long pacing). These times were chosen based on a pilot study by Kang et al. (2006). In this study, participants responded to the alpha-numeric questions at a rate of 2.5 seconds. The shortest feasible time between questions was 0.5. A maximum time between questions of 4 seconds was chosen as this value was maximum length of time provided for participants to respond in the Kang et al. (2006) study. A random number generator provided the initial question pacing and participants were allowed to adjust the pace throughout the first three test blocks. For example, if the participant decided they wanted more time between questions, they verbally informed the experimenter to increase the time between questions. Conversely, if the participant had a surplus time to respond, the experimenter reduced the pace following a verbal command from the participant. A computer program was used to present the questions to the participants at the pace specified. For consistency with traditional psychophysical studies, 3 test blocks were used to allow for pace adjustment (roughly 20.5 minutes). Participants completed4 more blocks at the selected pace from their adjustment session.

3.6.4 INDEPENDENT VARIABLES

Independent variables for this study included block (7 levels), test session (2 levels), and initial work pace (2 levels). Gender differences were also considered. Block consisted of the 7 question blocks within a single trial. Test session was defined as Adjustment Session (the first three blocks when participants were allowed to adjust question pacing) and Work Session (the final four blocks when participants worked at their selected question pace). Initial work pace was defined as short and long, as described above. The generation of the work pace categories is arbitrary, but based on the preliminary data that average response time was about 1.6 seconds (Kang et al, 2006).

3.6.5 DEPENDENT VARIABLES

Dependent variables for this study included thermal images of the face, subjective ratings, participant reaction time, and accuracy rate. The procedures used to collect data for this study was identical to those in study one. A brief summary is provided here.

Thermal Images

A MikronScan 7200V Thermal Camera (Mikron Infrared, Inc., Oakland, NJ) was used to collect facial thermal images. The camera was located in 45cm in front of the participant. Images were collected at a rate of 1Hz for the duration of the study. A baseline facial thermal image was collected prior to testing and following a 15 minute stabilization period, using the same procedure as study 1. Regions of Interest (ROIs) were superimposed over the forehead and nose (figure 3.6). Nose ROIs did not include either nostril as air flow during breathing may affect thermal readings. Forehead ROIs included the part of the face between the eyebrows and hairline (figure 3.6). This study collected maximum nose and forehead temperature in ROI because blood vessel that was affected by ANS should have a higher temperature than other area in ROI. Thermal readings were adjusted using a baseline image prior to statistical analysis (reading – baseline value).

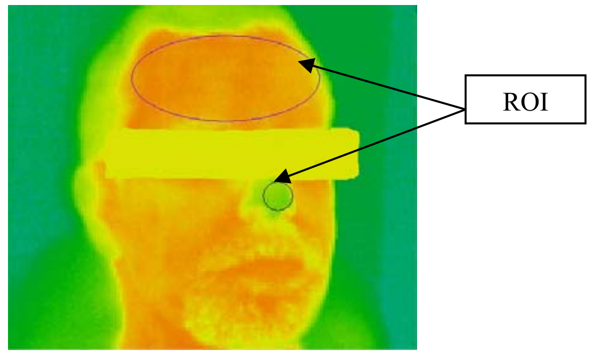


Figure 3.6

Thermal Image with ROI

Subjective Ratings of Mental Workload

Two subjective workload assessment tools, the Modified Cooper Harper Scale (MCH) and the Subjective Workload Assessment Technique (SWAT), were used in this study. Participants completed the subjective ratings during rest periods between blocks of questions. The MCH is a single dimensional technique that was developed to be used in cognitive and perceptual tasks (Appendix A). The SWAT is the three dimensional technique (time load, mental effort load, and overall workload) (Appendix B). This study used a modified version of the SWAT using three visual-analog scales (VAS). The VAS version of SWAT has been shown to be more sensitive in measuring moderate levels of mental workload (Luximon and Goonetilleke, 2001).Further details of these tools can be found in Wierwille and Casali (1983) and Reid and Nygren (1988) respectively.

Performance measures

Accuracy rate and reaction time was recorded. Accuracy rate was the percentage of questions answered correctly for each block. Reaction time was measured from the time of question presentation to the time that the participant responded using the keyboard.

3.6.6 PROCEDURE

Participants completed informed consent documents prior to any data collection. Demographic questionnaires were completed, followed by a 15 minute acclimation period. The baseline thermal images were collected followed by the first experimental session (which consists of 3 blocks). The second session consisted of 4 blocks, using a fixed response time gained by the first session after the first session. At the end of the second session, participants were compensated for their time.

3.6.7 DATA ANALYSIS

Appropriate descriptive statistics were generated for each dependent variable. A mixed factors ANOVA was used to determine differences between the dependent variables as a function of the independent variables (test session, block, work pace, and gender). Temperature at the key question and rate of change for each thermal image frame was analyzed. Tukey's HSD post hoc tests were conducted where appropriate. Correlations were calculated between each dependent variable. All findings were considered significant at an alpha of 0.05. Based on the data, additional analyses were conducted. Participants were grouped based on performance (high and low performers) and the same analyses described above were rerun to quantify differences in performance groups. Further details are provided below.

3.7 RESULTS

Descriptive statistics for each dependent variable is provided in Table 3.1. Blocks 1 through 3 constitute the adjustment session, while blocks 4 through 7 constitute the work session that participants were working at their defined "optimal pace". During the adjustment session, it was observed that, in general, adjusted nose temperature (ΔNT), accuracy, and reaction time (RT) increased, and subjective workload ratings (MCH and SWAT) and adjusted forehead temperature (ΔFT) decreased. RT, accuracy, and ΔFT remained constant, subjective workload ratings (MCH and SWAT) decreased slightly, and ΔNT varied for the work session in general.

Table 3.1

Descriptive statistics for the dependent variables (values are in mean (standard deviation))

Session	Block	ΔNT (^oC)	RT (sec)	Accuracy	MCH	SWAT	ΔFT (^oC)
Adj	1	-0.22	1.35	84.47	3.92	206.6	0.04
		(0.54)	(0.21)	(11.8)	(1.90)	(99.7)	(0.16)
	2	0.04	1.38	90.55	3.64	189.3	-0.01
		(0.66)	(0.21)	(5.86)	(1.72)	(98.31)	(0.20)
	3	0.05	1.37	91.86	3.25	176.9	-0.04
		(0.95)	(0.24)	(6.70)	(1.69)	(89.77)	(0.25)
Work	4	-0.23	1.38	92.55	3.21	163.5	-0.07
		(0.78)	(0.25)	(6.20)	(1.37)	(94.07)	(0.28)
	5	-0.16	1.38	93.50	3.10	160.3	-0.06
		(0.79)	(0.25)	(5.27)	(1.42)	(91.64)	(0.29)
	6	-0.28	1.39	93.19	3.03	150.7	-0.06
		(1.04)	(0.26)	(5.61)	(1.40)	(97.47)	(0.32)
	7	-0.58	1.37	92.62	2.96	144.3	-0.06
		(1.32)	(0.24)	(7.21)	(1.42)	(95.08)	(0.31)

Adj = Adjustment

3.7.1 OVERALL EXPERIMENT SESSION

All dependant variables were affected by block, except RT (Table 3.2). No gender differences or block by gender interaction effects were found. ΔNT was found to be similar across blocks 1 to 3, and block 7 was significantly lower than this grouping. Standard deviation in ΔNT (ΔNT S.D.) was found to be similar across block 1 to 6 and block 7 was smaller than blocks 3 and 4. ΔFT was found to be similar across block 1 to 3, and blocks 4 to 7, though differences were found between these two groupings. Accuracy for block 1 was significantly lower than the other blocks. SWAT and MCH ratings were similar for blocks 1 and 2, and for blocks 3 to 7, through there were differences between these groupings.

Table 3.2

Mixed factors ANOVA results for the overall experimental session

Dependent Variable	Block	Gender	Block * Gender
Accuracy	**< .0001**	0.5230	0.6258
RT	0.8924	0.9886	0.3850
ΔNT	**< .0001**	0.5927	0.8900
ΔFT	**0.0023**	0.7289	0.8479
SWAT	**< .0001**	0.1250	0.5931
MCH	**< .0001**	0.2819	0.8893
ΔNT S.D.	**0.0059**	0.4967	0.8739

Bolded values denote significant findings (p-value < 0.05)

Strong and significant correlations, greater than ±0.84, were found between accuracy and the other dependant variables, except ΔNT and ΔNT S.D. (Table 3.4). ΔNT was strongly and significantly correlated with ΔNT S.D. (-0.82). ΔFT was found to be strongly and significantly correlated, greater than ±0.91, with accuracy and subjective workload measures.

Because the Tukey groupings indicated trends that differentiated between the adjustment and work session, analyses between these two sessions were conducted.

Table 3.3

Tukey's post hoc comparisons for each task block for the overall experimental session

Dependant variable	Block	Mean	Group			Dependant variable	Block	Mean	Group	
Accuracy	1	84.47 %	A			ΔNT	1	-0.22 °C	A	
	2	90.55 %		B			2	0.04 °C	A	
	3	91.86 %		B			3	0.05 °C	A	
	4	92.55 %		B			4	-0.23 °C	A	B
	5	93.50 %		B			5	-0.16 °C	A	
	6	93.19 %		B			6	-0.28 °C	A	B
	7	92.62 %		B			7	-0.58 °C		B
SWAT	1	149.70	A			ΔNT S.D.	1	0.24	A	B
	2	122.90	A	B			2	0.19	A	B
	3	92.40	A	B	C		3	0.18	A	
	4	83.15		B	C		4	0.18	A	
	5	61.05		B	C		5	0.19	A	B
	6	53.80		B	C		6	0.24	A	B
	7	51.85			C		7	0.28		B
MCH	1	3.92	A			ΔFT	1	0.04 °C	A	
	2	3.64	A	B			2	-0.01 °C	A	B
	3	3.25		B	C		3	-0.04 °C	A	B
	4	3.21		B	C		4	-0.07 °C		B
	5	3.10		B	C		5	-0.06 °C		B
	6	3.03		B	C		6	-0.06 °C		B
	7	2.96			C		7	-0.06 °C		B

Table 3.4

Correlation coefficients for all significant correlations across all blocks

	ACC	RT	MCH	ΔFT	ΔNT	SWAT	ΔNT S.D.
ACC	1						
RT	**0.84**	1.00					
MCH	**-0.91**	-0.64	1.00				
ΔFT	**-0.96**	-0.74	**0.95**	1.00			
ΔNT	-0.12	0.06	0.44	0.27	1.00		
SWAT	**-0.87**	-0.65	**0.97**	**0.91**	0.59	1.00	
ΔNT S.D.	-0.20	-0.27	-0.13	0.13	**-0.82**	-0.28	1

Bolded values denote significant correlation coefficients (p-value < .05)

3.7.1.1 ADJUSTMENT SESSION ANALYSES IN OVERALL SESSION

For this session, ΔNT, ΔFT, accuracy, and MCH were significantly affected by block. No gender differences, or block by gender interaction effects, were found (Table 3.5). Block 1 was found to be lower than the other blocks in accuracy and ΔNT (Table 3.6). Block 1 for ΔFT and MCH were significantly different from block 3, while block 2 was similar to blocks 1 and 3.

Table 3.5

Mixed factors ANOVA results for the adjustment session

Dependent Variable	Block	Gender	Block * Gender
Accuracy	**< .0001**	0.4350	0.3549
RT	0.6433	0.9475	0.1360
ΔNT	**0.0168**	0.3304	0.9195
ΔFT	**0.0080**	0.4844	0.9668
SWAT	0.1230	0.2918	0.6051
MCH	**0.0447**	0.3712	0.8549
ΔNT S.D.	0.0937	0.6382	0.5303

Bolded values denote significant findings (p-value < 0.05)

Significant correlations were not found between most dependant variables, despite high correlation coefficients (Table 3.7). Strong and significant correlations were found between accuracy and ΔNT S.D. and between ΔFT and the SWAT.

Table 3.6

Tukey's post hoc comparisons for each task block in total adjustment session

Dependent variable	Block	Mean	Group		Dependent variable	Block	Mean	Group	
Accuracy	B1	84.47 %	A		ΔFT	B1	0.04 °C	A	
	B2	90.55 %		B		B2	-0.01 °C	A	B
	B3	91.86 %		B		B3	-0.04 °C		B
ΔNT	B1	-0.22 °C	A		MCH	B1	3.92	A	
	B2	0.04 °C		B		B2	3.64	A	B
	B3	0.05 °C		B		B3	3.25		B

Table 3.7

Correlation coefficients for the adjustment session for overall session

	ACC	RT	MCH	ΔFT	ΔNT	SWAT	ΔNT S.D.
ACC	1						
RT	0.84	1.00					
MCH	-0.90	-0.53	1.00				
ΔFT	-0.96	-0.68	0.98	1.00			
ΔNT	0.99	0.90	-0.83	-0.92	1.00		
SWAT	-0.96	-0.67	0.98	**0.99**	-0.92	1.00	
ΔNT S.D.	**-0.99**	-0.85	0.88	0.95	-0.99	0.95	1

Bolded values denote significant correlation coefficients (p-value < .05)

3.7.1.2 INITIAL PACE EFFECTS

Initial pace, slow or fast, was not found to affect any of the dependent variables (Table 3.8).

Table 3.8

Descriptive statistics and ANOVA results for initial pace effects

Dependant Variable	p-value	Mean (standard deviation)	
		Slow Initial Pace	Fast Initial Pace
Accuracy	0.2794	92.45 (4.81)	90.05 (9.67)
RT	0.7314	1.39 (0.20)	1.36 (0.27)
ΔNT	0.4508	-0.31 (0.57)	-0.08 (1.15)
ΔFT	0.6869	-0.02 (0.29)	-0.05 (0.24)
MCH	0.2481	3.0 (1.06)	3.6 (1.92)
SWAT	0.3094	153.9 (85.21)	186.6 (103.49)
ΔNT S.D.	0.2757	0.189 (0.14)	0.248 (0.19)

Adjustment time (AT) was defined as the number of blocks needed for participants to identify their self selected pace. Chosen reaction time (CRT) was defined as the self selected pacing of questions (time between adjacent questions). Initial pace difference did not affect the AT (p-value = 0.4611) or CRT (p–value= 0.6884), though on the average, participants beginning with a slower initial pace took longer to identify their self selected pace (1.85 blocks for slow initial pacing vs. 1.64 blocks for fast initial pacing) and the pace was quicker (2.28 seconds between questions for slow initial pacing vs. 2.35 seconds for fast initial pacing).

3.7.1.3 WORK SESSION ANALYSES IN OVERALL SESSION

ΔNT and ΔNT S.D. were affected by block (Table 3.9). No gender or gender by block interaction effects were found. For both ΔNT and ΔNT S.D., block 7 was significantly lower than the other blocks (Table 3.10). Blocks 4 to 6 were similar for ΔNT, while blocks 6 and 7 were similar for ΔNT S.D.

No significant correlations were found except between SWAT and MCH ratings (0.96) and between SWAT ratings and ΔNT S.D. (-0.99) (Table 3.11).

Table 3.9

Mixed factors AVNOVA results for the work session

Dependent Variable	Block	Gender	Block * Gender
Accuracy	0.4940	0.6740	0.8562
RT	0.7885	0.9851	0.4087
ΔNT	**0.0002**	0.7766	0.7312
ΔFT	0.9780	0.8706	0.2867
SWAT	0.0985	0.0845	0.6913
MCH	0.2329	0.2534	0.1378
ΔNT S.D.	**0.0004**	0.9292	0.7010

Bolded values indicate significant findings ($p<0.05$)

Table 3.10

Tukey's post hoc comparisons for the work session

Dependent variable	Block	Mean	Group		Dependent variable	Block	Mean	Group	
ΔNT	B4	-0.23 °C	A		ΔNT S.D.	B4	0.18	A	
	B5	-0.16 °C	A			B5	0.19	A	
	B6	-0.28 °C	A			B6	0.24	A	B
	B7	-0.58 °C		B		B7	0.28		B

Table 3.11

Correlation coefficients for work session

	ACC	RT	MCH	ΔFT	ΔNT	SWAT	ΔNT S.D.
ACC	1						
RT	0.57	1.00					
MCH	-0.07	0.27	1.00				
ΔFT	0.67	0.14	-0.79	1.00			
ΔNT	0.58	0.69	0.76	-0.21	1.00		
SWAT	0.17	0.31	**0.96**	-0.61	0.87	1.00	
ΔNT S.D.	-0.22	-0.33	-0.95	0.56	-0.89	**-0.99**	1

Bolded values denote significant correlation coefficients (p-value < .05)

3.7.2 HIGH PERFORMING PARTICIPANTS AND LOW PERFORMING PARTICIPANTS

After looking at trends in the data, it was observed that participants fell into one of two performance categories during the work session: low and high performers. Low performing participants had accuracies of less than 90%, where high performing participants had accuracies greater than 90%, and this difference was significant (p-value <0.0001) (Table 3.12). Given this trend, the analyses present above were re-run for these participants groups.

ΔNT and ΔNT S.D. were significantly different between the performance groups when considering the entire testing session (Table 3.12).High performers had significantly higher ΔNT and RT, and significantly lower ΔNT S.D. Though not significant, high performers had a longer RT and lower subjective workload ratings.

Table 3.12

ANOVA results between high and low performers for the entire testing session

Dependant Variable	p-value	Mean	
		High performers	Low performers
Accuracy (%)	< **0.0001**	94.08 (4.98)	84.17 (8.77)
RT (sec)	0.6193	1.39 (0.23)	1.34 (0.25)
ΔFT (^{o}C)	0.9106	-0.03 (0.26)	-0.04 (0.28)
ΔNT (^{o}C)	**0.0012**	0.09 (0.78)	-0.93 (0.80)
SWAT (cm)	0.4597	162.7 (86.79)	189.1 (114.45)
MCH	0.1724	3.07 (1.12)	3.8 (2.28)
ΔNT S.D.	< **0.0001**	0.15 (0.11)	0.37 (0.20)

Bolded values indicate significant findings (p-value<0.05)

As was done previously, the different sessions (adjustment and work period) were analyzed for high and low performers independently.

3.7.2.1 HIGH PERFORMER GROUP

In general, RT, and accuracy increased, while subjective mental workload ratings decreased across task blocks for high performers (Table 3.13). ΔNT was observed to increase through block 5, then decrease, while ΔFT remained relatively constant.

Table 3.13

Descriptive statistics for high performers (values are in mean (standard deviation))

Session	Block	ΔNT (°C)	RT (sec)	Accuracy (%)	MCH	SWAT (cm)	ΔFT (°C)
Adjustment	1	-0.07 (0.35)	1.37 (0.21)	86.87 (7.89)	3.80 (1.76)	207.3 (99.2)	0.04 (0.14)
	2	0.21 (0.67)	1.39 (0.21)	93.08 (2.89)	3.50 (1.46)	180.3 (92.3)	-0.01 (0.18)
	3	0.32 (0.97)	1.39 (0.23)	94.73 (4.32)	3.00 (1.16)	169.6 (81.5)	-0.04 (0.25)
Work	4	0.10 (0.62)	1.41 (0.23)	95.57 (2.62)	2.90 (0.64)	150.8 (76.5)	-0.07 (0.29)
	5	0.11 (0.73)	1.39 (0.24)	95.98 (2.12)	2.90 (0.64)	150.9 (76.0)	-0.06 (0.30)
	6	0.05 (0.96)	1.41 (0.27)	95.92 (2.21)	2.80 (0.61)	143.3 (87.9)	-0.06 (0.32)
	7	-0.07 (1.01)	1.38 (0.23)	96.45 (2.06)	2.63 (0.58)	136.7 (82.6)	-0.06 (0.31)

3.7.2.1.1 HIGH PERFORMER ADJUSTMENT SESSION

Block was found to significantly affect accuracy, ΔNT, and ΔNT S.D. No gender or gender by block interaction effects were found (Table 3.14).Accuracy in block 1 was significantly lower than blocks 2 and 3 (Table 3.15). Block 1 was also significantly lower than block 3 for ΔNT. Tukey's tests revealed no groupings despite a significant effect of block on ΔNT S.D.   Strong and significant correlations were found only between the subjective mental workload assessments and between SWAT ratings and ΔNT S.D. (Table 3.16).

Table 3.14

Mixed factors ANOVA results for the adjustment session for high performers

Dependent Variable	Block	Gender	Block * Gender
Accuracy	**< .0001**	0.9910	0.5711
RT	0.7983	0.5393	0.1561
ΔNT	**0.0098**	0.3767	0.5002
ΔFT	0.1448	0.6157	0.8673
SWAT	0.0970	0.7613	0.3422
MCH	0.0857	0.9047	0.9478
ΔNT S.D.	**0.0313**	0.0668	0.6501

Bolded values indicate significant findings (p-value < 0.05)

Table 3.15

Tukey's post hoc comparisons for the adjustment session for high performers.

Dependent variable	Block	Mean	Group		Dependent variable	Block	Mean	Group
Accuracy	B1	86.86 %	A		ΔNT S.D.	B1	0.22	A
	B2	93.08 %		B		B2	0.15	A
	B3	94.73 %		B		B3	0.14	A
ΔNT	B1	-0.07 °C	A					
	B2	0.20 °C	A	B				
	B3	0.31 °C		B				

Table 3.16

Correlation coefficients for high performers' adjustment session

	ACC	RT	MCH	FT	NT	SWAT	ΔNT S.D.
ACC	1.00						
RT	0.57	1.00					
MCH	-0.07	0.27	1.00				
ΔFT	0.67	0.14	-0.79	1.00			
ΔNT	0.58	0.69	0.76	-0.21	1.00		
SWAT	0.17	0.31	**0.96**	-0.61	0.87	1.00	
ΔNT S.D.	-0.22	-0.33	-0.95	0.56	-0.89	**-0.99**	1.00

Bolded values denote significant findings (p-value < 0.05).

3.7.2.1.2 HIGH PERFORMER WORK SESSION

There were no significant effects found for the high performers work session (Table 3.17). ΔNT was strongly and significantly correlated with subjective mental workload assessments and ΔNT S.D. (Table 3.18). A strong and significant correlation was also found between the subjective mental workload assessments.

Table 3.17

Mixed factors ANOVA results for high performers' work session

Dependent Variable	Block	Gender	Block * Gender
Accuracy	0.3342	0.6301	0.3419
RT	0.3297	0.5341	0.4335
ΔNT	0.0949	0.2038	0.1775
ΔFT	0.9668	0.9865	0.6046
SWAT	0.4127	0.4361	0.2145
MCH	0.3362	0.9020	0.0668
ΔNT S.D.	0.2390	0.1406	0.9675

Table 3.18

Correlation coefficients for high performers' work session

	ACC	RT	MCH	ΔFT	ΔNT	SWAT	ΔNT S.D.
ACC	1.00						
RT	-0.85	1.00					
MCH	-0.87	0.35	1.00				
ΔFT	0.53	0.80	-0.25	1.00			
ΔNT	-0.86	0.34	**0.99**	-0.20	1.00		
SWAT	-0.83	0.52	**0.98**	-0.32	**0.97**	1.00	
ΔNT S.D.	0.79	-0.46	**-0.97**	0.30	**-0.96**	**-0.99**	1.00

Bolded values denote significant correlation coefficients (p-value < 0.05).

3.7.2.2 LOW PERFORMER GROUP

In general, accuracy and RT increased and ΔNT and subjective mental workload assessments decreased across blocks (Table 3.19). ΔFT decreased from block 1 to 2, then remained constant for the remaining blocks.

Table 3.19

Descriptive statistics for low performers (values are in mean (standard deviation))

Session	Block	ΔNT (°C)	RT (sec)	Accuracy(%)	MCH	SWAT (cm)	ΔFT(°C)
Adjustment	1	-0.61 (0.76)	1.32 (0.23)	78.50 (17.66)	4.25 (2.31)	205.1 (107.8)	0.06 (0.21)
	2	-0.37 (0.41)	1.36 (0.24)	84.25 (6.80)	4.00 (2.32)	211.8 (115.3)	-0.02 (0.23)
	3	-0.61 (0.47)	1.33 (0.26)	84.70 (6.34)	3.88 (2.58)	195.1 (111.7)	-0.08 (0.28)
Work	4	-1.08 (0.45)	1.32 (0.29)	85.01 (6.24)	4.00 (2.26)	195.2 (128.9)	-0.07 (0.29)
	5	-0.84 (0.49)	1.36 (0.29)	87.32 (5.80)	3.63 (2.50)	183.7 (125.7)	-0.08 (0.29)
	6	-1.12 (0.76)	1.34 (0.23)	86.40 (5.83)	3.63 (2.44)	169.5 (122.7)	-0.08 (0.34)
	7	-1.89 (1.12)	1.37 (0.28)	83.06 (6.52)	3.75 (2.43)	163.3 (125.5)	-0.07 (0.35)

3.7.2.2.1 LOW PERFORMER ADJUSTMENT SESSION

No significant differences were found in the low performers' adjustment session (Table 3.20), and no significant correlations were found except between MCH ratings and ΔFT (Table 3.21).

Table 3.20

Mixed factors ANOVA results for low performers' adjustment session.

Dependent Variable	Block	Gender	Block * Gender
Accuracy	0.2265	0.7309	0.5823
RT	0.7541	0.2091	0.8384
ΔNT	0.3210	0.9784	0.1043
ΔFT	0.0699	0.6763	0.4458
SWAT	0.5021	0.2318	0.6714
MCH	0.5716	0.2813	0.9947
ΔNT S.D.	0.7425	0.0578	0.5708

Table 3.21

Correlation coefficients for adjustment session for low performers

	ACC	RT	MCH	ΔFT	ΔNT	SWAT	ΔNT S.D.
ACC	1.00						
RT	0.70	1.00					
MCH	-0.96	-0.49	1.00				
ΔFT	-0.94	-0.42	**0.99**	1.00			
ΔNT	0.45	0.94	-0.19	-0.12	1.00		
SWAT	-0.17	0.56	0.43	0.49	0.79	1.00	
ΔNT S.D.	0.11	0.78	0.15	0.23	0.93	0.95	1.00

Bolded values denote significant coefficients (p-value < 0.05).

3.7.2.2.2 LOW PERFORMER WORK SESSION

ΔNT and ΔNT S.D. were significantly affected by block (Table 3.22). No gender or gender by block interaction effects were found. Block 7 was significantly different from the other blocks for bothΔNT S.D. and ΔNT (Table 3.23). Further, block 6 for ΔNT was found to differ significantly from the other blocks, and block 4 was similar to block 5. ΔNT S.D. was similar across blocks 4 to 6. The only significant correlation was between SWAT ratings and ΔNT S.D. (Table 3.24).

Table 3.22

Mixed factors ANOVA results for low performers' work session

Dependent Variable	Block	Gender	Block * Gender
Accuracy	0.4940	0.6740	0.8562
RT	0.7885	0.9851	0.4087
ΔNT	**0.0002**	0.7766	0.7312
ΔFT	0.9780	0.8706	0.2867
SWAT	0.0985	0.0845	0.6913
MCH	0.2329	0.2534	0.1378
ΔNT S.D.	**0.0004**	0.9292	0.7010

Bolded values indicate significant findings (p-value < 0.05)

Table 3.23

Tukey's post hoc comparisons for low performers' work session

Dependent variable	Block	Mean	Group	Dependent variable	Block	Mean	Group
ΔNT	B4	-1.08	A	ΔNT S.D.	B4	0.31	A
	B5	-0.84	A		B5	0.36	A
	B6	-1.12	A		B6	0.48	B
	B7	-1.89	B		B7	0.57	C

Table 3.24

Correlation coefficients for work session

	ACC	RT	MCH	ΔFT	ΔNT	SWAT	ΔNT S.D.
ACC	1.00						
RT	-0.08	1.00					
MCH	-0.46	-0.68	1.00				
ΔFT	-0.90	0.21	0.52	1.00			
ΔNT	0.92	-0.38	-0.09	-0.78	1.00		
SWAT	0.37	-0.64	0.61	-0.14	0.69	1.00	
ΔNT S.D.	-0.56	0.57	-0.43	0.33	-0.83	**-0.97**	1.00

Bolded values denote significant correlation coefficients (p-value < .05)

A significant difference in the trend of ΔNT S.D. was found between the performance groups (p-value = 0.0032, Table 3.11), with high performers having a significantly lower ΔNT S.D. High performers ΔNT S.D. remained constants after block 1, while low performers had an increase in their ΔNT S.D. during work session (from block 4 to 7).

3.8 DISCUSSION

For the purposes of this discussion, the test session will be divided into the adjustment and work session for clarity, as significant differences between the adjustment and the work session were identified. Results for the different participant performance groups will be presented.

3.8.1 TEST HYPOTHESES

Hypothesis 1: Performance will be affected by block (i.e., performance will improve until reaching the optimal workload level).

In general, this hypothesis was not supported. Reaction time (RT) was not affected by block, which was indicated by participants answering questions at a constant rate. Accuracy was only found to be significantly affected by block in the high performer's adjustment session. Accuracy for block 1 was significantly lower than the other blocks, primarily due to a fast initial pacing, resulting in participants missing a number of questions. Further, requiring participants to solve the questions and adjust the question pacing simultaneously likely affected accuracy for this group. This resource sharing caused an increase in mental workload (Zakay, 1993; Hancock and Chignell, 1988; Chiles and Alluisi, 1979; Johannsen, 1979). Following the adjustment session, no performance differences were found. These findings are consistent with those in chapter 2 and others (Kang et al., 2006; Kang and Babski-Reeves, 2008).

Low performers did not have any performance differences between blocks. Likely this was due to their chosen question pacing (RT). The RT for low performers was lower than that of high performers, as well as their accuracy. This shortened RT may have lead to increased stress and fatigue, which have been found to affect mental workload (Johannse, 1979; Kang and Babski-Reeves, 2008; Wickens and Hollands, 2000;

Conrad, 1956; Reid and Nygren, 1988; Hendy et al., 1997; Hancock and Chignell, 1988; Zakay, 1993).

Hypothesis 2: Facial thermal readings will stabilize as the work level is adjusted toward optimal.

In general, there was support for this hypothesis. High performers' nose temperature followed what was found in chapter 2 and previous studies (Kang el al., 2006; Kang and Babski-Reeves, 2008), with change in ΔNT found in the adjustment session, but not in the work session. For low performers, the opposite was found (ΔNT differed in the work session, not the adjustment session).

What was interesting was that when considering all participants, ΔFT was significantly affected by block in the adjustment session. This finding differed from previous research where forehead temperature was not affected by mental workload (Stall 1964; Veltoman and Vos, 2005, Kang and Babski-Reeves, 2008; Kang et al, 2006). As much of the mental processing for this type of task (problem solving, memory, and judgment) occurs in the frontal lobe, the need for blood flow to this area during the adjustment period may have resulted in this finding (Guyton and Hall, 2006; Shimamura et al, 1995). Also, the multi-tasking (answering questions and adjusting question pacing) may have resulted in this finding. Previous research has indicated that multi-tasking requires higher mental workload by increasing task requirements (Kaber et al., 2001).

This difference was only present for blocks 1 and 3. This was because participants selected their work pace (RT) before the end of the second block, therefore, this multi-tasking was not present (or minimally present) during the third block.

Hypothesis 3: Thermal readings, performance, and subjective workload measures will be correlated.

Very few correlations were found in this study. When considering all participants as a group, most of the correlations related to subjective measures and a thermal measure (though the specific thermal measure differed). These trends were consistent when considering high and low performers independently. During the adjustment session, the lack of correlations is somewhat expected due to the task requirements. However, it was expected that during the work session, when the workload was supposed to be "optimal" that the measures would be correlated. The lack of correlation may have been due to the inability of the psychophysical method to allow for the identification of the true "optimal" workload level. This is discussed further in the next session.

Low performers had only one significant correlation between MCH ratings and forehead temperature. Because these participants were not performing as well, it is reasonable to assume that these participants were doing more mental activities in the frontal lobe, as mentioned previously, resulting in the relationship with forehead temperatures (Guyton and Hall, 2006; Shimamura et al, 1995). Furthermore, participant differences may have introduced significant variation that overshadowed any underlying relationship. High performers had significant correlations between nose temperature and the subjective workload tools, consistent with chapter 2 findings.

As found in previous studies, subjective mental workload measures were strongly correlated with each other for high performers in both sessions, indicating that between two subjective workload measures differences were minimal (Kang et al., 2006; Kang and Babski-Reeves, 2008). Contrary to previous studies, however, was that subjective measures were not correlated with thermal and performance measures.

No correlation with performance measures was found for either participant performance group, which was contrary to chapter 2 or previous findings (Kang et al., 2006; Kang and Babski-Reeves, 2008). This is likely due to the ceiling effect for performance with changes in subjective ratings and thermal readings. Another possible reason for the discrepancy in the finding with previous research is the objective. In the previous studies, the goal of the research was to assess learning. Therefore, as participants became more familiar with the task, their reaction time increased. This study was not interested in learning since that should have already occurred. Rather, this study was attempting to determine if participants could choose a work pacing that would optimize workload levels. Participants in this study learned the pacing and adjusted their RT to this fixed pacing rate. The presence of a fixed pacing rate may also have affected the results. In the previous studies, the pacing of the questions was based on the speed of the participants, that is, questions were presented immediately following the response to the previous question. This means that pacing was somewhat variable. In this study, the pacing of the questions occurred at specific intervals and the participants were aware of this. Participants likely adjusted their responses to fit within this window and to allow time for rest.

Hypothesis 4: Participants will be able to identify an optimal question pacing to minimize workload.

A psychophysical approach was used to allow participants to identify the "optimal" question pacing level. However, it seems that they identified a level more appropriate for a short term task (about 5-7 blocks length task) rather than an 8 hour task duration. This hypothesis was supported by considering the results of chapter 2 in conjunction with these results. Average RT for study 1 was 1.8 seconds, with a shortest block average of 1.63 seconds, while average RT for this study was 1.37 seconds. This finding indicated that the participants overestimated their optimal workload level as their RT in this study was much shorter than the RT of the first study. In addition, RT was not different between stages, and likely was the result of learning the question pacing, as described in the previous discussion.

Most participants identified their self-selected work pace before the completion of block 2 (Figure 3.7). Participants, regardless of performance, took 1.75 blocks to self select their optimal question pacing. Both accuracy and ΔNT was found to increase over the adjustment session (Figure 3.7), which as similar to previous findings (Kang et al., 2006; Kang and Babski-Reeves, 2008). Therefore, participants' mental resources should no longer have been divided between answering the questions and evaluating the question pacing (Wickends and Holland, 2000). Accuracy indicated that participants in block 2 performed at a similar level with blocks 3 through 7, indicating that they reached their best performance level in block 2 followed by a speed accuracy trade-off (Conrad, 1956). Likely, participants were still evaluating question pacing though they made no further changes to the pacing.

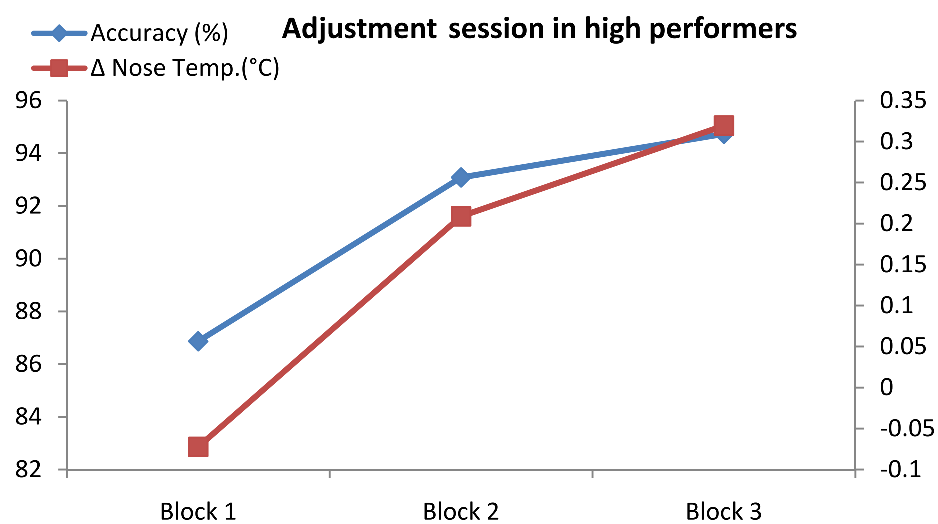


Figure 3.7

Thermography trend of adjustment session for high performers

For high performers' work session, ΔNT and ΔNT S.D. showed that the mental workload level was similar between blocks. This indicated that high performers found a workload level similar to their "optimal" level for short period work enabling them to reach a high accuracy rate (90%). The chosen workload level may not have been suitable for long term work because outside factors, such as fatigue and stress, have been found to increase mental workload level (Ross et al, 1975; Kahneman and Tversky, 1973; Andre and Wickens, 1995; Kang and Babski-Reeves, 2008). Also, those factors were not considered when participants determined their subjective ratings, as ratings decreased over the sessions. These findings were more pronounced in low performers as evidenced in Figure 3.8.

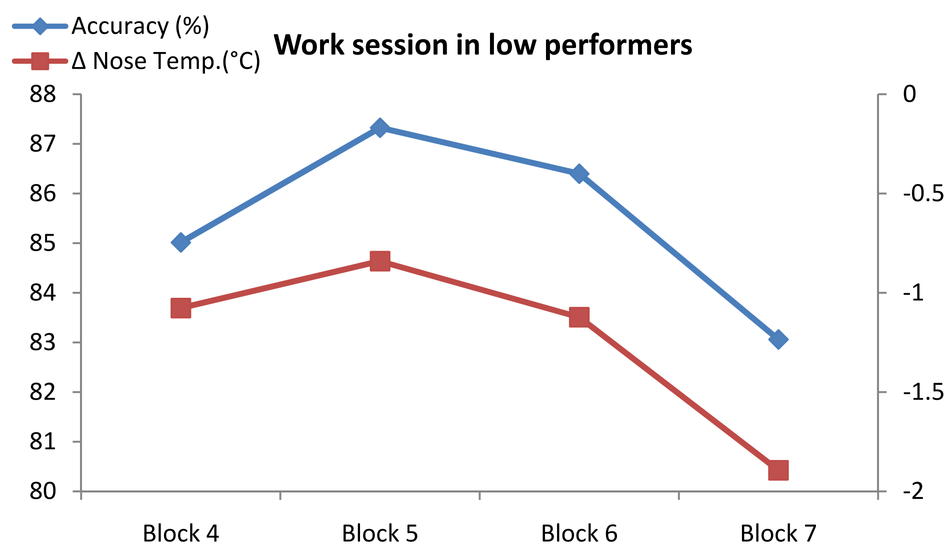


Figure 3.8

Thermography trend of evaluation session in low performers

Although all measures did not show a difference between blocks in high performers' work session in statistical results, thermography found a decreasing trend after block 5, indicating that mental workload was increasing. This study did not provided sufficient time to investigate whether performance would be affected by stress or fatigue, particularly for high performers, as indicated by thermograhpy. However, low performers in the work session found that accuracy decreased after reaching its' plateau. This is similar to the findings in chapter 2 and Kang and Babski-Reeves (2008). The study could, therefore, assume that high performers would experience a decrease in accuracy after block 7 because of boredom and stress.

All findings indicated that participants could not reach a skill-based high performance level because they overestimated their subjective optimal workload level (Rasmussen, 1983, 1986, 1993). Operators at optimal workload level can achieve high levels of skill-based performance level because they have lower mental workload, allowing them to process information automatically (Tsang and Vidulich, 2006). However, the participants operated at a rule-based performance level because they miscalculated the task when solving questions. They needed more training time to memorize, or at times become familiarized, with the simple alphabet arithmetic task within their subjective RT so that they could perform the task without hesitation at an expert level.

Hypothesis 5: Gender will not affect any dependent variable.

No gender differences were found for any session or performance group. This finding was consistent with previous research (Kang el al, 2006; Kang and Babski-Reeves, 2008).

3.8.2 WORKLOAD LEVEL OF HIGH PERFORMERS AND LOW PERFORMERS

Subjective workload measures (MCH and SWAT) were not found to differ between the performance groups, indicating that participants experienced similar mental workload levels regardless of performance. However, ΔNT and ΔNT S.D. showed that low performers experienced significantly higher mental workload, and that these participants failed to find their optimal workload level for this task. Further, these participants had an increasing trend in their workload level (Figure 3.9), possibly to be confounded by stress and fatigue (Kang and Babski-Reeves, 2008; Krueger, 1989).

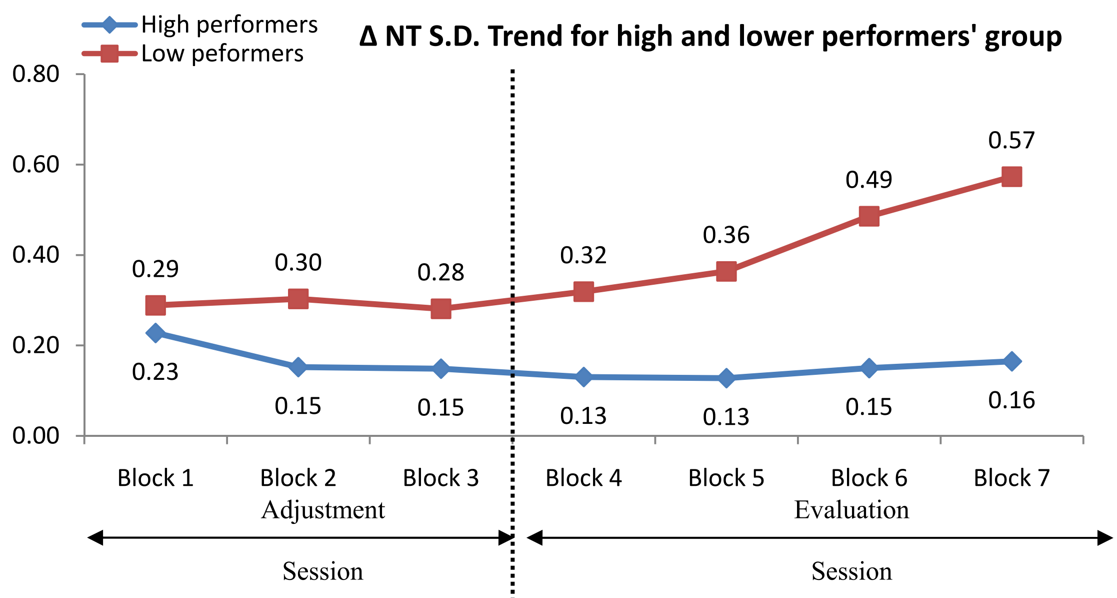


Figure 3.9

ΔNT Standard Deviation trend for the performance groups

High performers' ΔNT approached baseline after block 1, while low performers ΔNT was never higher than their baseline for the study (Figure 3.10). Findings for the high performers were similar to what was found in chapter 2, with ΔNT exceeding the baseline after learning (block 1) and during skill mastery (block 2 and 3). Low performers needed additional training to achieve a higher performance level. High performers did not need to train with the chosen RT because they chose similar optimal RT, although the chosen RT was not applicable to a long term task.

Thermography (ΔNT) demonstrated an ability to show profiles, in addition to averages, within blocks while performance and subjective measure show only one value of a single task block. The trend graphs are located in Appendices F and G. These figures showed differences between high performers and low performers. ΔNT for high performers remained constant within each block, indicated by smaller ΔNTS.D., resulting in good performance, as has been found previously (Kang and Babski-Reeves, 2008).Low performers exhibited a very sharp decreasing trend within each block, indicating their chosen work pace was non-optimal even for a short task. With these trend graphs, researchers can observe human mental workload fluctuation and distinguish whether participants find their optimal workload level.

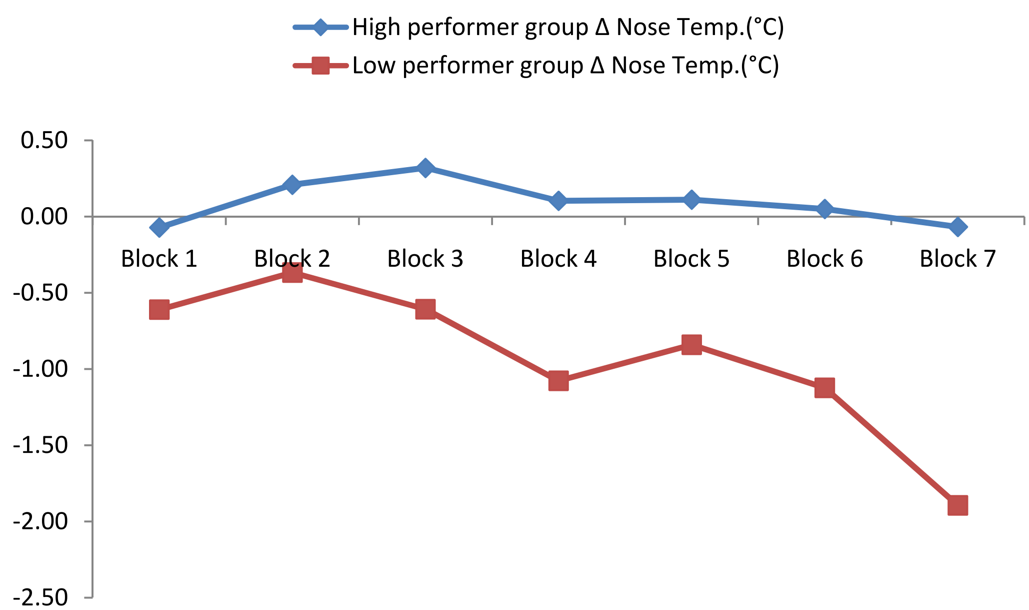


Figure 3.10

Trend of ΔNT in high and low performers

Subjective workload ratings were found decrease across the task blocks for both high and low performers. Previous discussion above posited that subjective ratings were not in line with true subjective ratings. Two theories can explain this finding. “Belief perseverance” states that if people decided or believed something previously, it will be difficult to change their mind even in the face of disconfirming evidence (Ross et al, 1975). The “anchoring and adjustment heuristic” states that people try to estimate or decide based on what they know previously (Kahneman and Tversky, 1973). Participants may have rated their mental workload wrongly because they believed they were becoming familiar with the task, which has been found in other research (Andre and Wickens, 1995; Kahneman and Tversky, 1973; Ross et al, 1975). Participants also may not have perceived the resulting stress and fatigue associated with task completion, leading to a mismatch between subjective workload ratings and other workload measures (O’Donnell and Eggemeier, 1986).

3.8.3 USE OF PSYCHOPHYSICS TO DETERMINE OPTIMAL WORKLOAD LEVELS

Findings in this study indicate that participants overestimated their work level. This is a consistent finding in other psychophysical studies (Colle and Reid, 1998; Mital, 1983; Colle and Reid, 2005). As previously explained in the “belief perseverance” theory, participants did not want to change their optimal workload level although they had one more block to adjust their subjective optimal RT. However, thermal readings in block 7 were lower than the other block, indicating increasing mental workload. This finding indicates that participants overestimated their optimal workload with psychophysical estimation methods. Subjective workload ratings were not sensitive to this finding however.

While it appears that there may be some promise to using a psychophysical type of approach to determining optimal workload levels, there are some concerns. First, typical psychophysics is used primarily for physical tasks, where it may be easier to "feel" the effects of the work level for prolonged periods of time. Correlations (0.84) between RT and accuracy in overall session also showed that performers increased their accuracy rate when they spent more time (RT) to answer the questions. High performers also used more time to answer the questions than lower performance on the average, indicating that RT was related with accuracy rate. Therefore, adjustment times may need to be increased. Also, in tasks such as this one when there is no feedback on performance, it may be difficult for participants to gauge at what level they need to work at to meet a minimum performance standard. The use of performance feedback may be useful in reducing the adjustment time.

3.9 FUTURE WORK

This study used the psychophysical estimation method for cognitive task estimation. Further research is needed to fully understand the use of this, or any estimation method, in finding optimal workload levels for cognitive tasks. Also, research is needed to investigate the effect of multiple tasks on ΔNT, particularly during the adjustment period. Thermography needs to be investigated under a longer experiment design to test an overestimating effect on participants for all dependant variables, particularly high performers.

3.10 CONCLUSION

This study demonstrated that thermography is able to distinguish between near optimal and non-optimal workload levels. Trends in thermal data can be used to identify workload levels and may be used to assist individuals in adjusting the workload to improve performance. Also, thermography explained mental workload fluctuations during workload adjustment periods more precisely than other mental workload measures (subjective and performance measures). Findings from this study further support the use of thermography as a valid, objective and non-invasive workload measure for mental tasks.

3.11 REFERENCES

Alvager, T., Anderson, E., French, V.A., Putman, G., & Shi, L. (1999). An artificial neural network study of the relationship between arousal, task difficulty and learning. *IJCNN '99. International Joint Conference on Neural Networks*, 5. 3618 – 3621.

Ayoub, M.M., & Dempsey P.G. (1999). The psychophysical approach to manual materials handling task design. *Ergonomics*, 42, 17-31.

Borg, G. (1990). Psychophysical scaling with applications in physical work and the perception of exertion. *Scandinavian Journal of Work Environment & Health. 16*, 55-58.

Chalder, T., Berelowits, G., Pawlikowska, T., Watts, L., Wessely,S., Wright, D., & Wallace, E.P. (1993). Development of a fatigue scale. *Journal of Psychosomatic Research*, 37, 147-153.

Chiles, W.D., & Alluisi, E.A. (1979).On the specification of operator or occupational workload with performance-measurement methods. *Human Factors*, 21, 515-528.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences(2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Colle, H.A., Amell, J.R., Ewry, M.E., & Jenkins, M.L. (1988). Capacity equivalence curves: a double trade-off curve method for equating task performance. *Human Factors*, 30, 645-656.

Colle, H.A. & Reid, G.B. (1998). Contest effects in subjective mental workload ratings. *Human Factors*, 40, 591-600.

Colle, H.A. & Reid, G.B. (2005). Estimating a mental workload redline in a simulated air-to-ground combat mission. *The International Journal of Aviation Psychology*, 15(4), 303-319.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial intelligence*, 16, 555-575.

Conrad, R. (1956). The timing of signals in skill. *Journal of Experimental Psychology,* 51, 365-370.

Gaillard, A.W.K. (1993). Comparing the concepts of mental load and stress. *Ergonomics*. 36, 991-1005.

Gawron, V.J. (2000). *Human Performance Measures Handbook*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. BoŠ, L. Kaufman, & J. Thomas (Eds.), *Handbook of Perception and Human Performance : Volume II. Cognitive Processes and Performance* (pp. 41-49). New York, NJ: Wiley.

Grant, K.A., D.J. Habes, & V. Putz-Anderson (1994).Psychophysical and EMG correlates of force exertion in manual work. *International Journal of Industrial Ergonomics,*13, 31-39.

Guyton, A.C., & Hall, J.E. (2006). *Text book of medical physiology (11th).* Philadelphia, PA: Elsevier Inc.

Hancock, P.A., & Chignell, M.H. (1988). Mental workload dynamics in adaptive interface design. *IEEE Transactions on Systems, Man and Cybernetics*, 18, 647-658.

Hancock, P. A., & Meshkati, N. (1988). *Human Mental Workload*. Elsevier Science: Amsterdam.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Handcock & N. Meshkati, (Eds.), *Human Mental Workload* (pp. 139-178). Amsterdam: North-Holland.

Hendy, K.C., Liao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information processing load. *Human Factors* 39, 30-47.

Hollnagel, E. (2006). Take analysis: why, what, and how. G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics (3rd ed.)(pp.* 373-383). Hoboken, NJ: John Wiley& Sons, Inc.

Jex, H. R. (1988). Measuring mental workload: Problems, progress, and promises, In P.A. Hancock & N.Meshkati, (Eds.), *Human Mental Workload* (pp. 5-39). New York: North-Holland.

Johannsen, G. (1979). Workload and workload measurement. In N, Moray (Ed.), *Mental Workload: Its Theory and Measurement* (pp. 3-12). New York: Plenum.

Kaber, D.B., & Riley, J.R. (1999). Adaptive Automation of a Dynamic Control Task Based on Secondary Task Workload Measurement. *International Journal of Cognitive Ergonomics*, 3, 169-187.

Kaber, D. B., Riley, J. M., Tan K., & Endsley, M. R. (2001). On the Design of Adaptive Automation for Complex Systems.*International Journal of Cognitive Ergonomics*, 5, 37 – 57.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction, *Psychology Review*, 80, 237-251.

Kang, J., McGinley, J., McFadyen, G., & Babski-Reeves, K. (2006). Determining learning level and effective training times using thermograhy. *Proceedings of 25th Army Science Conference*, Orlando, FL.

Kang, J. and Babski-Reeves, K. (2008). Detecting mental workload fluctuation during learning a novel task by thermography. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual meeting* , September 22-26, 2008, New York, NY.

Karwowski, W., & Yates, J.W.(1984). The effects of time in the psychophysical study of the maximum acceptable amounts of liquid lifted by females. *In Proceedings of the Human Factors Society's 28th Annual Meeting*, San Antonio, Texas, pp. 586–590.

Karwowski, W., & Yates, J.W. (1986). Reliability of the psychophysical approach to manual lifting of liquids by females.*Ergonomics,***29**, 237–248.

Kataoka, H., Kano, H., Yoshida, H., Saijo, A., Yasuda, M., & Osumi, M. (1998). Development of a skin temperature measuring system for non-contact stress evaluation. *Proceedings of the 20th Annual International conference of the IEEE engineering in Medicine and Biology Society*, 20 (2), 940-943.

Legg, S.J., & W.S. Myles (1981). Maximum acceptable repetitive lifting workloads for an 8-hour work-day using psychophysical and subjective rating methods. *Ergonomics, 24*, 907-916.

Logan, G.D., & Klapp, S.T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity?. *Journal of Experimental Psychology: Learning Memory, and Cognition*, 17, 179-195.

Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D. Linton, P.M., Wierwille, W.W., Zaklad, A.L., Bittner, A.C., & Wherry, R.J. (1989). *Operator workload: comprehensive review and evaluation of operator workload methodologies* (Technical Report 851). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Meister, D. (1976). *Behavioral Foundations of System Development.* New York: Wiley.

Mital, A. (1983). The psychophysical approach in manual lifting--a verification study.*Human Factors*, 25, 485-491.

Moray, N. (1979). *Mental Workload: Its Theory and Measurement*.New York: Plenum.

Moray, N. (1988). Mental workload since 1979. In D.J. Oborne (Ed.), *International Reviews of Ergonomics*: Vol. 2. *Current trends in human factors research and pratice (International reviews of ergonomics)* (pp. 123-150). London: Taylor & Francis.

Naemura, A., Tsuda, K., & Suzuki, N. (1993). Effects of loud noise on nasal skin temperature. *The Japanese Journal of Psychology*, 62, 51-54.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. BoŠ , L. Kaufman, & J. Thomas (Eds.), *Handbook of Perception and Human Performance : Volume II. Cognitive Processes and Performance* (pp. 42.1-42.9). New York: Wiley.

PUTZ-ANDERSON, V., & Grant, K. A. (1995). Perceived exertion as a function of physical effort, In S. L. Gordon, S. J. Blair and L. J. Fine (Eds), *Repetitive Motion Disorders of the Upper Extremity* (Rosemont: American Academy of Orthopaedic Surgeons), 61.

Rasmussen, J. (1983). Skills, rules, knowledge: Signals, signs, and symbols and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13, 257-267.

Rasmussen, J (1986). *Information processing and Human-machine interaction: An approach to cognitive engineering*. New York: Elsevier.

Rasmussen, J (1993). Deciding and doing: Decision making in natural contexts, In G.Klein, J, Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 158-171). Norwood, NJ: Ablex.

Reid, G.B. & Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock & N. Meshkati. (Eds). *Human Mental Workload* (pp. 185-213). Amsterdam: North-Holland.

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm, *Journal of Personality and Social Psychology,* 32, 880-892.

Scerbo, M. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman, & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp37-64). Mahwah, NJ: Lawrence Erlbaum.

Selcon, S.J. (1990). Decision support in the cockpit. Probably a good thing?. *Proceedings of the 34th annual meeting of the Human Factors and Ergonomics Society*, 46-50, Santa Monica, CA.

Shimamura, P., Jurica, P.J., Mangels, J.A., Gershberg, F.B., & Knight, R.T. (1995). Susceptibility to memory interference effect following frontal lobe damage: findings from tests of paired-associate learning. *Journal of Cognitive Neuroscience*, 7, 144-152.

Snook, S.H. (1978). The design of manual handling tasks. *Ergonomics*, 21, 963-985.

Tasng, P.S., & Vidulich M.A. (2006). Mental workload and situation awareness. In G. Salvendy (Ed.). *Handbook of Human Factors and Ergonomics (3rd ed.)* (pp. 234-268). New Jersey: John Wiley & Sons.

Thompson, L.A., Williams, k.l., L'Esperance, P.R., & Cornelius, J. (2001). Context-dependent memory under stressful conditions: The case of skydiving, *Human Factors,* 43, 611-619.

Veltman, H.J.A., & Vos, W.K. (2005). A methodology for non-intrusive mental workload measurement in simulated and on-road driving. *Proceedings of the 11th International conference on Human-computer interaction*, Las Vegas, NV.

Wickens, C.D., Gordon, S.E., & Liu, Y. (1998). *An Introduction to Human Factors Engineering.* New York: Addison Wesley Longman, Inc.

Wickens, C.D., & Hollands, J.G. (2000). Engineering Psychology and Human Performance *(3rd ed.).* New Jersey: Prentice_Hall.

Wierwille, W.W., & Casali, J.G. (1983). A validated rating scale for global mental workload measurement applications. *Proceedings of the Human Factors Society 27th Annual Meeting*, Norfork, VA, 129-133.

Yerkes, R.M., & Dodson, J.D., (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.

Young, M.S., & Stanton, N.A. (2001). Mental workload: theory, measurement, and application. In W. Karwowski (Ed.), *International Encyclopedia of Ergonomics and Human Factors* (pp. 507-509). New York: Taylor and Francis.

Zakay, D. (1993). Impact of time perception processes. In O. Svenson, & A.J. Maule (Eds.), *Time Pressure and Stress in Human Judgment and Decision Making* (pp. 59-72). New York, NY: Plenum Press.

CHAPTER 4

ASSESSING TRAINING TIME AND MENTAL WORKLOAD DURING COMPLEX TASK PERFORMANCE USING THERMOGRAPHY

4.1 ABSTRACT

Objective: The objective of this study was to investigate the efficacy of using thermography to quantify the relationship between human mental workload and facial temperature changes in a complex training/learning environment.

Methods: Twenty eight participants, 14 experienced and 14 novice drivers, completed 6 simulated driving stages. Each stage had city, rural, and highway scenarios. Changes in nose and forehead temperature, driving performance and two subjective mental workload ratings (MCH ratings and SWAT) were collected.

Results: Thermography and subjective workload measures showed that mental workload levels were similar between experienced and novice drivers, although experienced drivers performed better than novice drivers. Strong and significant correlations were, however, found among thermography, subjective workload measures, and performance measures in novice drivers' data.

Conclusion: Thermography measures mental workload during learning as non-invasive and objective measurement and is more practical than current subjective mental

workload measures in a complex training/ learning environment. Also, thermography has a capability to estimate a sufficient training time for a complex task although the training times are variable across tasks.

4.2 INTRODUCTION

The role of training has become increasingly more important to improve product quality and human safety in modern environments. The primary goal of training is to simplify the acquisition of required skills and knowledge for a specific task (Swezey and Pearlsten, 2001). Training can be described as a trainee's knowledge, skill (e.g., trainee competence), and attitude (e.g., working with confidence) about the learned task. Research has shown that training leads to improved learned task performance (Salas et al., 2006).

Efficient training results in minimized training time and cost while allowing for the achievement of acceptable performance levels (Wickens and Hollands, 2000). It is estimated that US organizations spend between $52.4 and $200 billion annually on training (Salas and Cannon-Bowers, 2001; Galvin, 2002). Additionally, the number of hours spent in training has increased from 24 hours in 2001 to 28 hours in 2002 (American Society for Training and Development, 2003). Providing appropriate minimum training times for skill mastery and concept retention can reduce costs. However, little literature exists for identifying appropriate minimum training times.

Traditional training uses a fixed average training time estimated by past experience or expert observation (Joyce, 1999), and does not take into account individual differences among trainees. As individuals are known to learn at varying rates, using standardized training times leads to potential problems. A trainee not provided with sufficient training time may make unexpected errors and may have lower than acceptable task performance. Conversely, trainees who require less training time than is scheduled will become bored and unnecessary training costs will be expended. Basing training time on performance measures can be useful in identifying minimum, required training times.

During training, mental workload levels are elevated due to learning; though workload levels decline as learning occurs. Mental workload is described as the physiological and mental demands for performing a task (FAA, 2005) based on an operator's subjective perception and work experience (Hart and Staveland, 1988). These definitions suggest that there is a relationship between mental workload, human performance, and physiological activity. Therefore, measuring mental workload using physiological measures during training may be an effective way of identifying when sufficient training has occurred. This research will investigate the utility of using thermography as a mental workload assessment tool to prescribe sufficient training time for driving simulator tasks.

4.3 BACKGROUND

4.3.1 TRAINING AND MENTAL WORKLOAD

Learning requires cognitive activity and training facilitates the learning process, particularly in the early stages (Wickens et al., 1998). Cognitive activities include perception (of displays, the environment, feedback, etc.) and working memory (transferring material to long-term memory, feedback processing, evaluating alternatives, etc.), among others. Thus, many training and instructional techniques interfere with the learning process because they require high demands on working memory initially (Wickens and Hollands, 2000). As learning occurs, however, training contributes to reduced cognitive workload by reducing loads imposed on working memory (Sanders and McCormick, 1993). Experts have been found to have greater memory than novices performing the same task because of training, resulting in experts having improved task performance, even when exposed to random task events (Vicente, 1992; Ye and Salvendy, 1994; Wickens and Hollands, 2000).

Trainee's cognitive ability (mental ability) is an important factor influencing training outcomes (Salas et al., 2006). Knowledge acquisition about a job is a strong determinant of training success (e.g., Ree et al., 1995; Ree and Earles, 1991; Randel et al., 1992; Colquitt et al., 2000), promoting self-efficacy (operator's own ability) and skill acquisition (e.g., Hunter, 1986). These findings indicate that cognitive ability is important for learning and training and suggest the purpose of measuring mental workload.

Ultimate impacts of training and learning are reduced mental workload levels during job task performance, particularly in complex work environments. For example, Collet et al. (2003) found that bus drivers using automated docking systems initially experience higher mental workload levels than bus drivers using a manual docking process. However, over time (practice and learning) mental workload levels for automated docking bus drivers were reduced below their manual bus driver counterparts. The training process imposes severe demand on the learner's cognitive resources, and this increases the learner's mental workload because the working memory demand is coincident with perception demand. If working memory and perception are overloaded, effective learning cannot occur (Wickens et al., 1988). Training techniques such as adaptive, guided, and part-task training have been developed to reduce a trainee's mental workload.

There are several training methodologies that can be employed to minimize mental workload levels and loads on trainee cognitive abilities (e.g, adaptive training, guided training, and part task training). Adaptive training systematically varies difficulty levels during training according to the individual trainee's skill level (Johnson and Haygood, 1984; Wickens et al., 1998). This allows for trainee's to experience varying scenarios that require progressively more simultaneous cognitive processes, thereby exposing them to more real world scenarios for a specific task. Guided training does not allow errors during training by imposing constraints on the trainee to prevent mistakes. Preventing errors during training prevents trainees from learning inappropriate behaviors and/or responses thereby increasing task performance (Wickens et al., 1998; Catrambone and Carroll, 1987). Part-task training takes a task and breaks that task into a series of simple tasks (Wickens et al, 1998; Wickens and Hollands, 2000; Wightman and Lintern, 1985). Trainees then learn how to perform each subtask individually, and then integrate the subtasks into an entire task. If previous subtasks are repeated with the introduction of each subsequent subtask, then integration during training is occurring continuously, allowing for early subtasks to become more automatic.

Feedback on trainee performance is required to determine when learning is occurring. When skills become automated during training, cognitive resource loads will be reduced (Fisk et al., 1987; Rogers et al., 1999; Schneider, 1985; Wickens and Hollands, 2000; Collet et al, 2003). Feedback can also identify knowledge or skill gaps of trainees that can be addressed (Salas et al., 2006). Sufficient training is hypothesized to occur when improvements in trainee performance are no longer observed.

4.3.2 MENTAL WORKLOAD MEASURE AND THERMOGRAPHY

Currently, several mental workload assessment techniques exist, though no single mental workload measure is consistently employed across studies. Therefore, continued research into inclusive, innovative mental workload assessment tools exist. Physiological mental workload assessment tools are attractive since they do not interfere with job task performance. Previous research has indicated that thermography might be a viable non-contact mental workload assessment technique (Kang et al., 2006; Genno et al., 1997 [a, b];

Veltman and Vos, 2005; Green and Shellenberger, 1991; Trujillo, 1998). Increases in mental workload levels are indicated by decreases in thermal readings of the nose. Further studies on using thermography to quantify mental workload are needed, however.

4.3.3 SIMULATION AND DRIVER EDUCATION

Training assistance technology has provided innovative training methods such as computer-based instruction and simulation (Salas, 2006). Simulators have been a preferred training method in business, education, and military (Jacobs and Dempsey, 1993). Driving simulators were initially used to evaluate fatigue and age effects (Salas, 2006). Now, driving simulator research is extending its territory as a popular method of driver training, safety training, and driver assessment (e.g., Fisher et al., 2002; Roenker et al., 2003). Two primary advantages of using driving simulators are reduced maintenance cost and increased safety.

A benefit of simulators is that they allow trainees to make mistakes and receive feedback without catastrophic consequences. Simulators can expose trainees to more realistic situations than other training methods, with reduced risk (no risk to humans). Fidelity of driving simulators has been evolving with the use of emerging high-speed computer and electronic technology, which supports the collection of highly accurate data and provides realism to participants. Human factors researchers concur that a modern advanced driving simulator has many benefits over similar real world or on-road driving research (Godley, 2002). High fidelity training simulations provide highly realistic and accurate training environment (Wickens et al., 1998) with relatively less cost (figure 4.1). Other advantages to using driving simulators include experimental control, efficiency, and ease of data collection (Nilsson, 1993).

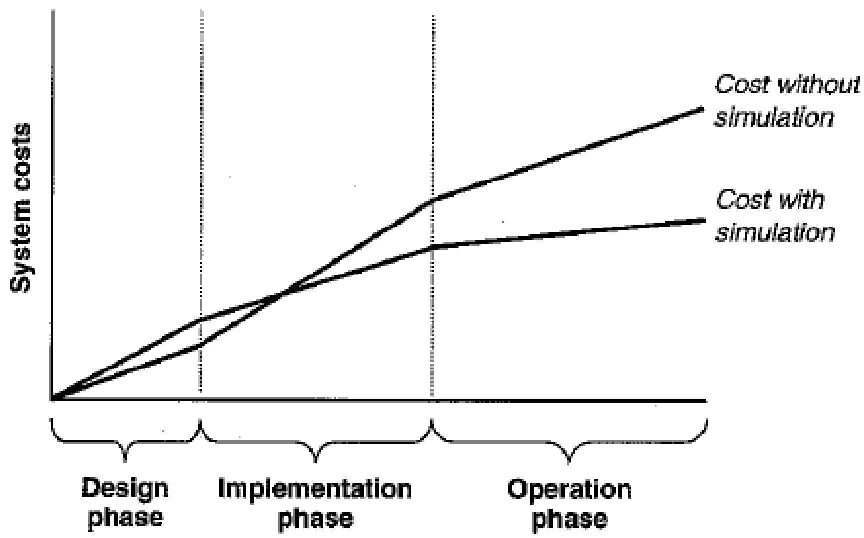


Figure 4.1

Comparison of cumulative system costs with and without simulation (Harrell et al., 2004).

Researchers agree that using driving simulators can improve driving skills (e.g. Roenker et al., 2003; Lee et al., 2003). Lee et al. (2003) confirmed the high transferability of observations between simulated driving and on-road assessment. Furthermore, simulator training has the potential to reduce accident rates because simulator training can provide various opportunities for drivers to practice specific situations (scenarios) that are unsafe (e.g., emergency stopping or overtaking other cars) as well as normal driving situations (Ehlert and Rothkrantz, 2001).

Driving simulators also allow for testing of a driver's unsafe or risky driving behavior (Allen et al., 1990; Carsten et al., 1997; Alicandri, 1994; Fraser et al., 1994; Van der Winsum, 1996; Desmond and Matthews, 1997; Ellingrod et al., 1997; Van der Winsum and Brouwer, 1997). Driving simulator systems provide an easily programmable scenario definition language for designing driving courses and providing performance measurements. Since most driving simulators have an automatic data collection function, the vehicle's status is monitored and measured easily during research and training. For these reasons, driving simulators have been using to train novice drivers.

4.3.4 NOVICE DRIVERS AND DRIVING SIMULATION

Novice drivers are confronted with many difficult situations and new tasks, requiring increased use of mental resources. Also, novices take more time and effort than experienced drivers to react in the same situation (Gregerson, 1996). Novice drivers are more likely to die in a crash than more experienced drivers (Pradhan et al., 2005). The accident rate for novices is 9.3 fatal crashes per 100 million vehicle miles; however, experienced drivers between 45-54 years age have an accident rate of 1.4 fatal crashes per 100 million vehicle miles (Insurance Institute for Highway safety, 2003).

McKnight and McKnight (2003) indicate that the novice drivers' lack of experience is the major cause of crashes (42.7% of the crashes). Gregerson (1996) indicated that a deficit of experience is the cause for up to 70% of all crashes among young drivers. However, novice drivers do not have any chance to experience or learn how to deal with risky or hazardous situations in traditional driving training programs because of safety concerns. Deficiency of driving experiences may affect a novice drivers' mental ability to apply the high-level cognitive skills for driving in very demanding scenarios (Mayhew and Simpson, 1995; Ranney, 1994).

The primary benefit of simulation training is that novice drivers can have various experiences interacting with cognitively complex driving scenarios without risk, and these experiences contribute to improvements in their situational awareness, risk assessment, and decision making in critical situations (Allen et al., 2003). This function of driving simulator training contributes to a reduction in the potential of novice drivers' crash rate and an improvement in their driving skill in risky driving situations (Fisher et al., 2002). Literature concludes that a driving simulator training program is a useful and validated training method to train novice drivers.

4.4 OBJECTIVES

Researchers have demonstrated that the training process reduces trainee's mental workload. There is also a relationship between mental workload and human facial temperature, particularly nose temperature. However, there are no evaluation measures to show how mental workload fluctuates during the training process. This study employs thermography as a non-contact and objective measure to detect the change in mental workload during training.

The objective of this study is to introduce a methodology that demonstrates a measurable correlation between facial temperature changes and the human mental workload in a complex training environment. Further, this study demonstrates a method for determining sufficient training time by using changes in facial temperature to monitor mental workload. This new method will contribute to reduce training time and cost. It will also contribute to determining an optimal training time for each individual. In addition, this study will identify a time that is required until an experienced driver is adapted to a new environment (driving simulator). This adaptation time also will be considered as a training time for the experienced drivers to master a driving skill in a simulator environment.

4.5 HYPOTHESES

The specific hypotheses to be tested included:

1. The process of training and practicing to obtain a driving task skill will reduce mental workload over time.
2. Facial temperature readings will have a relationship with a learner's performance and subjective mental workload measures.
3. A learner's performance will have a relationship with mental workload.

4. There will be no gender differences in mental workload or performance.
5. Novice and experienced drivers will perform differently.
6. Novice and experienced drivers will improve their performance until they have mastered the skill in the simulation environment. After this point, their performance will improve slightly or remain constant.
7. Facial temperature readings will remain constant after learning has occurred.

4.6 METHODOLOGY

4.6.1 EXPERIMENTAL DESIGN

A mixed subject design was used to assess the effects of stage, a within subjects factor (6 levels) and driver group, a between subjects factor (2 levels), on thermography readings of the face, subjective workload assessment readings, and training times. This experiment used two groups, experienced drivers (more than 3 years of driving experience) and novice drivers (less than 1 year of driving experience). Each group completed 6 driving simulator stages, each lasting for approximately 15 minutes. A 5-minute rest period between stages was provided. Performance measures included brake reaction time, percentage of vehicle lane deviation, violations of traffic regulations, lost control, and collision rate. Exposures to test stages were determined using a Balanced Latin Square.

4.6.2 PARTICIPANTS

Twenty eight participants, 14 males and 14 females, completed experimental protocols. The number of participants is based on an effect size of 1.0, α=0.5, which results in a power of 72% (Cohen, 1988). Participants were defined as experienced drivers, drivers with more than 3 years of experience, and novice drivers, drivers with less than 1 year of experience. Participants ranged in age from 14 to 17 years for novice drivers and from 18 to 25 for experience drivers. Each group consisted of equal numbers of males and females. All participants were required to have at least 20/20 or corrected to 20/20 vision (based on self report) and not be colorblind.

4.6.3 TASK DESCRIPTION

A driving simulator provided a three-phase scenario to each participant: a city phase with two parallel parking events, a rural phase with S-curves, and a highway phase. The phases were randomly presented to participants within a stage. The experiment consisted of six, 15-minute stages, with a 5-minute rest period between stages. HyperDriver software (DriveSafety, Inc., Murray, UT) was used to provide a driving simulation environment. The simulator included a Dodge Neon car seat with manual controls for adjustment, steering wheel, CRT (Cathode Ray Tube) monitor (19-inch) for presenting driving scenarios, dashboard, turn signal, and brake and gas pedal. The driving simulator utilizes various built-in driving landscapes produced from the HyperDrive software, around which traffic flows can be constructed. Two speakers located on each side of the driver to provide realistic sounds timed with the driving scenario.

Environment Description

Scenarios consisted of a city, rural, and highway setting presented randomly to the participants. The city setting depicted metropolitan constructions and roadways. The driving course was a circuit formed by four right-turn or left-turn intersections with a traffic light or sign at each of the turns. Various vehicles were moving on the opposite side of road and parked vehicles were on the sides near the curbs. The roadway was a 2-lane road with a speed limit of 30 mph. Participants were required to complete two parallel parking events during this portion of the scenario. The highway setting included a divided 4-lane highway with moderate traffic. Participants were instructed to drive at a speed limit of 65 mph and the driver was encouraged to keep his/her speed at least 50 mph in this stage. Participants were required to make lane changes and adjust speed based on traffic patterns. The rural setting depicted typical rural-area constructions, such as farmhouses and a grassy plain, and provided a low traffic condition. The roadway was a 2-lane road with double solid yellow stripes in the middle. The speed limit was restricted to 55 mph, and participants were encouraged to at least 40 mph. The rural roadway included a 1:10 downward slope with two, sharp S-curves at random points in the scenario. Each stage included randomly placed, unexpected events in which a pedestrian or animal walked across the road unexpectedly, requiring participants to decelerate or make a complete stop to avoid collision.

4.6.4 INDEPENDENT VARIABLES

The independent variables of this study included driving stage (6 stages), driving group, and gender. Driving groups consisted of experienced and novice drivers. Experienced group performance data was used as a criterion to test whether novice drivers learned driving skills with the driving simulator. Each performance was measured through stages including city, highway, and rural scenarios.

4.6.5 DEPENDENT VARIABLES

Dependent variables for this study included facial thermal readings, two subjective workload ratings, and driving performance measures. Thermal readings were collected continuously during stages. Subjective workload ratings were collected following each scenario in each stage. Performance measures were described below.

Thermography Readings

A MikronScan 7200V Thermal Camera (Mikron Infrared, Inc., Oakland, NJ) was used to measure changes in facial temperature. The MikronScan 7200V is a non-contact, high sensitivity infrared radiometer. This fully-radiometric camera can simultaneously record high-definition, 14-bit thermal images with digital visual images. The range of measurable temperatures of the camera is 0°C to 500°C (32.0°F to932.0°F) with the sensitivity/NETD of 0.08°C. The camera was located 45cm in front of the participant, though it did not impede their visual sight. Images were collected at a rate of 1Hz for the duration of the study. A baseline image was collected prior to data analysis following a 15-minute acclimation period and used to obtain delta facial thermographic readings for analysis. Delta facial thermographic readings were obtained subtracting the baseline reading from the recorded temperature for each frame. For recorded temperature, this study collected nose and forehead temperature. Laboratory temperature was controlled to avoid environmental effects on the facial temperature. Circle-shaped ROIs (Regions of Interest) over the forehead area and nose area (Figure 4.2) were used to develop assessment regions for data analysis. ROI for nose did not include either nostril. Forehead ROIs included the part of the face between the eyebrows and hairline (figure 4.2).This study collected maximum nose and forehead temperature in ROI because blood vessel that was affected by ANS should have a higher temperature than other area in ROI.

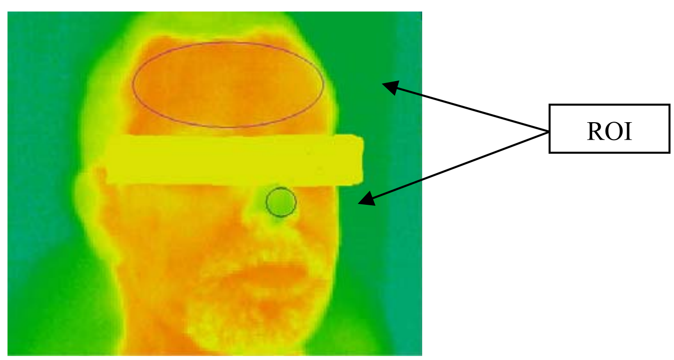


Figure 4.2

Thermal Image with ROI

Driving simulator performance data

The simulator automatically collected data such as lost control, lane deviation of vehicle, vehicle speed (average speed, over-speed), brake reaction time, and collision (descriptions provided below). For this study the following data was collected manually or through the simulator: brake reaction time (Richter and Hyman, 1974; Johansson and Rumar, 1971; Green, 2000), vehicle lane deviation (Dorn and Barker, 2005; Summala et al., 1996), violation of traffic regulations (Lee et al., 2003) (such as not observing speed limits, stop signs, over-speed, traffic lights, etc.), and collision rate. Performance data collection was synchronized with the thermal camera recording interval time (one second).

- Brake reaction time: Total brake reaction time can be split into perception/perceive-to-move time and movement time (Johansson and Rumar, 1971; Liebermann et al 1995; Green, 2000).
    - Perceive-to-move time was defined as the time from the onset/presentation of the stopping stimulus until the foot began to release the accelerator.
    - Movement time was defined as the time for a driver to move the foot from the accelerator to the brake in the braking situation.
- Lost control: Driver can’t control their vehicle as what they want.
- Lane deviation: Driver can’t maintain their vehicle in the middle of the lane and the vehicle is out of lane more than 2/3 of the vehicle body.
- Average speed: speed is averaged in each stage.

- Over-speed: Occurrences of speeding in each stage.
- Collision: Occurrences of collision in each stage.

4.6.6 PROCEDURE

Participants were asked to complete informed consent documents prior to data collection. Participants completed a demographic and medical history questionnaire (see *Appendix C)* and pre-exposure simulator sickness questionnaire (see *Appendix D)* to determine eligibility. Following a 15-minute acclimation period, a baseline thermal reading was taken. Participants adjusted the simulator to their desired settings, and testing began. A total of six, 15-minute scenarios separated by 5-minute rest periods were completed. During the rest periods, participants completed subjective workload ratings and a post-exposure simulator sickness questionnaire (see *Appendix E).* At the completion of testing, participants were compensated for their time.

4.6.7 DATA ANALYSIS

A mixed factor ANOVA was used to assess the effects of stage, driver group and gender on facial thermal readings, performance measures (participant brake reaction time, vehicle lane deviation, lost control, violation of traffic regulations, and collision), and subjective workload ratings. Tukey's HSD post hoc tests were employed to determine whether there were significant differences between factors levels where appropriate. Correlations between dependent variables were also computed. Because of trends identified in the data during initial analyses, additional analyses were conducted as depicted in Figure 4.3. The same analyses described above were conducted to assess changes with the session phases (adjustment and work) and within participant performance categories (high performers and low performers). Details for these categories are presented below.

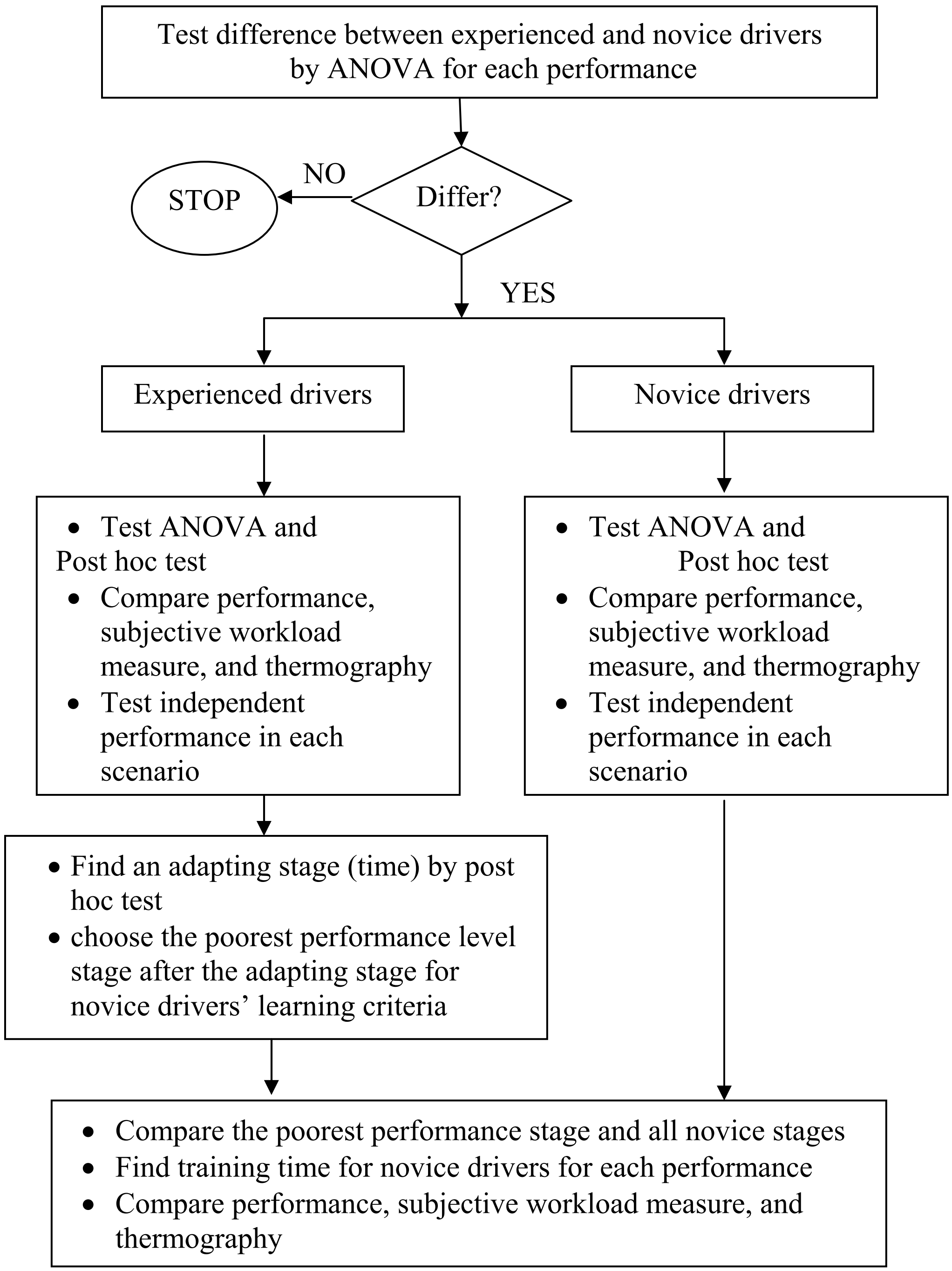


Figure 4.3

Hierarchy diagram for analysis

Further, this study investigated two different driver's groups; experienced and novice drivers. Therefore, experienced group performance data was used as a criterion to test whether novice drivers learned driving skills with the driving simulator. That is for each performance measure, the poorest performance block for the experienced driver group was compared to each block of the novices. When no differences between the driver and experience group are found, this implies that the novices had learned the skill.

Lastly, each performance was measured through stages including city, highway, and rural scenarios. Three different scenarios, however, had different specific independent events due to their environment limitations such as a parking event for city, an S-curve for rural, and lane deviation for rural and highway scenarios. Therefore, performances during these scenario specific events were considered separately from other performance measures. All analyses were performed in SAS 9.1 and all findings were considered significant at an alpha level of 0.05.

4.7 RESULTS

Experience level was found to affect only performance measures (Table 4.1 and 4.2). For all performance measures, experienced participants were found to have superior performance than novices. No gender differences were found for any of the dependent variables (Table 4.1).

Three experiences by gender interactions were found: ΔNT, average speed, and parking duration. No consistent patterns were identified in the Tukey test for ΔNT and average speed (Table 4.3). Parking duration for each gender was found to be at a similar level when participants were in the same experience group. Experienced male drivers had shorter parking duration times than novice male and female drivers (Figure 4.4).

Table 4.4 shows means and standard deviations for the significantly performance differences between experienced and novice drivers' result in table 4.1. Novice drivers showed significantly higher violation performance rate than experienced drives in collision rate, lost control rates, and lane deviation (Table 4.4). Experienced divers showed significantly shorter performance time, with higher accuracy, in perception time, movement time, total brake time, parking time and failure, and S-curve performance.

Table 4.1

Mixed factors ANOVA results for the overall test session

Measurement	Dependent variables	Experience	Gender	Experience * Gender
Thermography	ΔNT	0.3347	0.1392	**0.0263**
	ΔFT	0.4527	0.0767	0.3119
	ΔNT S.D.	0.2997	0.4172	0.7332
Subjective measures	SWAT	0.1172	0.4736	0.1712
	MCH	0.3572	0.7506	0.1978
Performance measures	Perception time	< **.0001**	0.7790	0.2806
	Movement time	**0.0112**	0.4311	0.8800
	Total brake time	< **.0001**	0.5227	0.5903
	Lost control	**0.0003**	0.6616	0.8838
	Collision	**0.0001**	0.3851	0.1618
	Over-speed	0.1831	0.1359	0.5099
	Average speed	0.2351	0.4673	**0.0447**

Bolded values denote significant effects

Table 4.2

Mixed factor ANOVA results for independent performance measures in each scenario for both driver groups

Scenario	Dependent variables	Experience	Gender	Experience * Gender
City	Blinker	0.1512	0.8834	0.1512
	Parking duration	**0.0002**	0.7800	**0.0232**
	Parking failure	**0.0001**	0.7096	0.0530
Highway	Lane deviation	< **0.0001**	0.9704	0.8528
Rural	S-curve	**0.0396**	0.8684	0.8684
	Lane deviation	< **0.0001**	0.8692	0.6220

Bolded values denote significant effects

Table 4.3

Tukey's post hoc comparisons for interactions between experience and gender for overall stage and independent events in each scenario

Dependent variable	Interaction	Mean	Group	Dependent variable	Interaction	Mean	Group
ΔNT	Exp. Female	-1.88	A	Parking duration	Exp. Female	56.97	A C
	Exp. Male	0.75	A		Exp. Male	35.98	A
	Nov.Female	0.38	A		Nov.Female	72.13	B C
	Nov. Male	-0.18	A		Nov. Male	88.74	B
Average speed	Exp. Female	40.83	A				
	Exp. Male	43.01	A				
	Nov.Female	41.51	A				
	Nov. Male	40.46	A				

Exp. = Experienced and Nov. = Novice

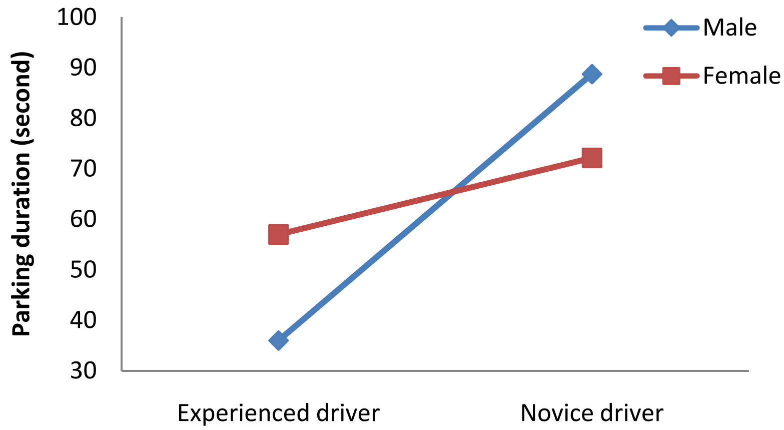


Figure 4.4

Interaction effect between experience and gender on parking duration

Table 4.4

Descriptive statistics for the significantly different performance between experienced and novice driver results in Table 4.1

Dependent variables	Mean (S.D.)	
(scenario)	Experienced driver	Novice driver
Perception time	0.278 (0.156)	0.740 (0.372)
Movement time	0.405 (0.197)	0.555 (0.309)
Total brake time	0.682 (0.256)	1.279 (0.551)
Lost control	0.095 (0.176)	0.325 (0.365)
Collision	0.531 (0.527)	1.317 (1.118)
Parking duration (City)	46.476 (25.202)	80.436 (36.038)
Parking failure (City)	0.452 (0.609)	1.178 (0.823)
Lane deviation (Highway)	0.130 (0.339)	2.261 (1.729)
S-curve (Rural)	0.190 (0.395)	0.345 (0.478)
Lane deviation (Rural)	0.261 (0.517)	1.142 (0.880)

Dependent variables (Table 4.3), affected by drivers' experience level significantly, were tested to reveal when novice reached a similar skill level with experienced drivers. First, this study observed experienced drivers' performance to investigate whether they became familiar with a new simulated driving environment (Adaptation Time). The poorest performance level for the experienced drivers was used to investigate whether novice drivers learned the skill.

4.7.1 EXPERIENCED DRIVERS

ΔNT showed an increasing trend across stages, while ΔFT and ΔNT S.D. showed a general decreasing trend (Table 4.5). Collision and lost control rates also decreased across stages, while braking performance (including perception, movement, and total brake time) remained relatively constant. Subjective workload measures showed a continuous decreasing trend across stages.

No gender differences were found for any of the dependant variables, with the exception of ΔNT (Table 4.6). Female ΔNT (-1.88^{o}C) was found to be significantly lower than male ΔNT (0.75 ^{o}C).

Stage differences were found for ΔNT S.D., MCH ratings, SWAT ratings, collision, lost control rates, and total brake time (Table 4.6). Stage 2 was found to be significantly higher than stage 6 ΔNT S.D., though no other differences between stages were found (Table 4.7). This same trend was observed for total brake time (Table 4.7). Collision rates for stage 1 were found to be higher than stage 6, and were found to be similar for stages 1 to 5 and stages 2 to 6. For lost control rates, stage 2 was higher than stage 5, while a similar level was found for stages 1 to 4 and for stages 3 to 6. Subjective workload measures were found to be similar for stages 1 and 2, stages 2 and 3, and across stages 4 to 6.

Table 4.5

Descriptive statistics for the dependent variables for the experienced drivers for overall (values are in mean (standard deviation))

Stage	ΔNT (^{o}C)	ΔFT (^{o}C)	ΔNT S.D.	MCH	SWAT (cm)
1	-0.85 (2.56)	0.16 (0.29)	0.30 (0.13)	3.73 (1.26)	196.2 (65.44)
2	-0.43 (2.37)	0.02 (0.38)	0.35 (0.21)	3.35 (1.14)	166.5 (62.22)
3	-0.40 (2.38)	0.03 (0.39)	0.26 (0.09)	2.92 (0.83)	142.1 (55.87)
4	-0.68 (2.83)	-0.05 (0.39)	0.24 (0.09)	2.59 (1.02)	122.6 (72.25)
5	-0.48 (2.80)	-0.11 (0.55)	0.25 (0.12)	2.38 (0.85)	111.4 (67.09)
6	-0.53 (2.42)	-0.01 (0.45)	0.23 (0.09)	2.14 (0.72)	98.4 (60.35)

Stage	Collision (rate of occurrence)	Lost control (rate of occurrence)	Perception time (sec)	Movement time (sec)	Total brake time (sec)
1	0.85 (0.50)	0.11 (0.16)	0.30 (0.13)	0.34 (0.21)	0.64 (0.25)
2	0.71 (0.65)	0.21 (0.28)	0.33 (0.21)	0.52 (0.16)	0.86 (0.26)
3	0.38 (0.41)	0.04 (0.12)	0.26 (0.13)	0.38 (0.27)	0.64 (0.26)
4	0.47 (0.38)	0.11 (0.16)	0.30 (0.18)	0.42 (0.13)	0.73 (0.21)
5	0.52 (0.67)	0.02 (0.08)	0.26 (0.13)	0.40 (0.21)	0.67 (0.30)
6	0.23 (0.24)	0.04 (0.12)	0.18 (0.07)	0.34 (0.09)	0.53 (0.10)

Table 4.6

Mixed factor ANOVA results for the experienced drivers for overall session

Dependent variable	Gender	Stage	Stage * Gender
ΔNT (^{o}C)	**0.0368**	0.7360	0.7569
ΔFT (^{o}C)	0.5461	0.1080	**0.0024**
ΔNT S.D.	0.7674	**0.0287**	0.2525
MCH	0.2070	< **0.0001**	**0.0370**
SWAT	0.1725	< **0.0001**	0.4334
Collision	0.5456	**0.0136**	0.9889
Lost control	0.7257	**0.0229**	0.0884
Perception time	0.4066	0.0832	0.2785
Movement time	0.2832	0.1489	0.9548
Total brake time	0.9030	**0.0101**	0.4956

Bolded values denote significant effects

Table 4.7

Tukey's post hoc results for experienced drivers overall session for stage

Dependent variable	Stage	Mean	Group				Dependent variable	Stage	Mean	Group	
ΔNT S.D.	1	0.30	A	B			Collision	1	0.85	A	
	2	0.35	A					2	0.71	A	B
	3	0.26	A	B				3	0.38	A	B
	4	0.24	A	B				4	0.47	A	B
	5	0.25	A	B				5	0.52	A	B
	6	0.23		B				6	0.23		B
MCH	1	3.73	A				Lost control	1	0.11	A	B
	2	3.35	A	B				2	0.21	A	
	3	2.92		B	C			3	0.04	A	B
	4	2.59			C	D		4	0.11	A	B
	5	2.38			C	D		5	0.02		B
	6	2.14				D		6	0.04	A	B
SWAT	1	196.2	A				Total brake time	1	0.64	A	B
	2	166.5	A	B				2	0.86	A	
	3	142.1		B	C			3	0.64	A	B
	4	122.6		B	C			4	0.72	A	B
	5	111.4			C			5	0.67	A	B
	6	98.4			C			6	0.53		B

Significant stage by gender effects were found for ΔFT (Figure 4.5) and MCH ratings (Figure 4.6). Male, stage 5 ΔFT was found to be significantly lower than male, stage 1 and 2 ΔFT. No other differences were found (Table 4.8). Female, stage 1 MCH ratings were found to be higher than all other stages except female, stage 2 ratings. Female stage 1 MCH ratings were also found to be significantly higher than male, stages 4 to 6 MCH ratings (Table 4.8).

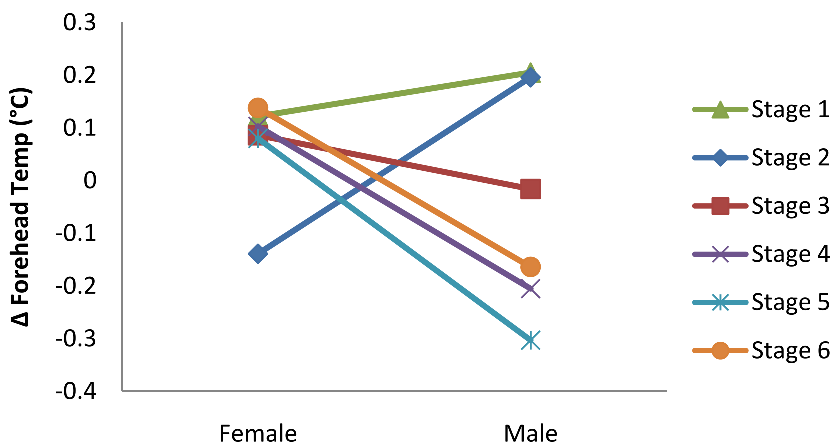


Figure 4.5

Interaction effect between stage and gender on ΔFT

for experienced driver for overall

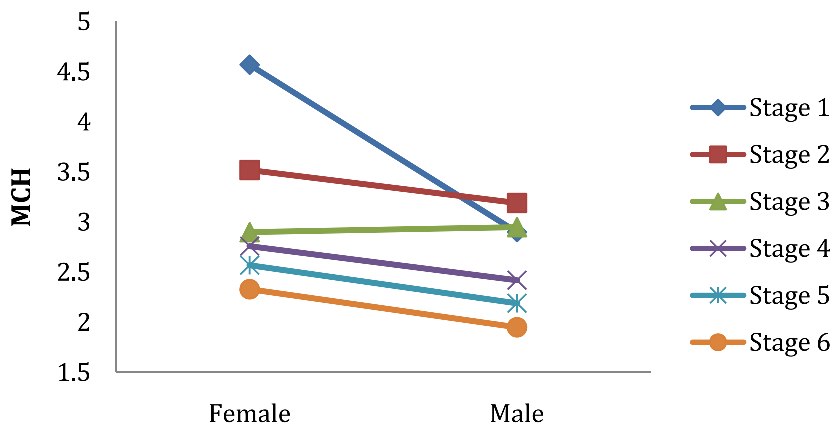


Figure 4.6

Interaction effect between stage and gender on MCH ratings

for experienced drivers for overall

Table 4.8

Tukey's post hoc results for the experience by gender interaction in the overall session for experienced drivers

Dependent variable	Interaction	Mean	Group		Dependent variable	interaction	Mean	Group	
MCH	Female Stage 1	4.57	A		ΔFT	Female stage 1	0.12	A	B
	Female Stage 2	3.52	A	B		Female Stage 2	-0.13	A	B
	Female Stage 3	2.90		B		Female Stage 3	0.08	A	B
	Female Stage 4	2.76		B		Female Stage 4	0.10	A	B
	Female Stage 5	2.57		B		Female Stage 5	0.08	A	B
	Female Stage 6	2.33		B		Female Stage 6	0.13	A	B
	Male Stage 1	2.90	A	B		Male Stage 1	0.20	A	
	Male Stage 2	3.19	A	B		Male Stage 2	0.19	A	
	Male Stage 3	2.95	A	B		Male Stage 3	-0.01	A	B
	Male Stage 4	2.42		B		Male Stage 4	-0.20	A	B
	Male Stage 5	2.19		B		Male Stage 5	-0.30		B
	Male Stage 6	1.95		B		Male Stage 6	-0.16	A	B

Strong, and signficant, correlations were found between ΔFT and subjective measures, indicating ΔFT and subjective workload measures increased similarly (Table 4.7). However, ΔNT was not found to correlate significantly with the other dependent variables for the experienced driver group. No performance measures were found to correlate significantly with thermal readings or subjective ratings, except collision which correlated strongly with SWAT ratings.

Table 4.9

Correlation coefficients for the overall session for experienced drivers

	ΔFT	ΔNT	ΔNT S.D.	MCH	SWAT	Collision	Lost control	Perception time	Movement time	Total brake time
ΔFT	1									
ΔNT	-0.52	1								
ΔNT S.D.	0.50	0.01	1							
MCH	**0.83**	-0.40	0.80	1						
SWAT	**0.85**	-0.45	0.78	**0.99**	1					
Collision	0.58	-0.51	0.78	**0.78**	**0.88**	1				
LC	0.39	-0.15	**0.82**	0.64	0.61	0.63	1			
PC	0.27	-0.21	0.73	0.73	0.70	0.81	0.74	1		
MT	-0.21	0.41	0.65	0.22	0.16	0.30	0.74	0.66	1	
TBT	-0.02	0.17	0.74	0.47	0.42	0.55	0.80	**0.87**	**0.94**	1

Bolded values denote significant effects (LC = lost control, PT = perception time, MT = movement time, Total brake time = TBT)

4.7.1.1 INDEPENDENT EVENTS IN EACH SCENARIO FOR EXPERIENCED DRIVERS

Each stage had three scenarios and each scenario has specific events where performance on the events were analyzed: parking event for city, S-curve for the rural scenario, and lane deviation for the rural and highway scenarios. For the independent events, this study analyzed drivers' performance measures with each scenario's thermal readings and subjective workload measures.

City scenario

Gender differences were found for ΔNT, parking duration, and parking failure (Table 4.10). Females had lower ΔNT, longer parking durations, and more parking failures than males in the city environment (Table 4.11).

Stage differences were found for ΔNT S.D., subjective workload, parking duration, and parking and failure. ΔNT S.D. for stage 2 differed from stages 3 to 6, though no other differences were found (Table 4.12). Stage 1 was significantly higher than the other stages for SWAT ratings, MCH ratings, and parking duration. Parking duration in stage 1 was similar to stage 2, while parking duration was at similar level for stages 3 to 6. Parking failures for stage 1 was similar to stage 4, while parking failures for stages 2 to 6 were similar.

A stage by gender interaction effect was found for ΔFT (Table 4.10). Male, stage 1 ΔFT was significantly higher than male, stage 4 and 5 ΔFT (Figure 4.7). Male, stage 2

ΔFT was significantly higher than male stage 5 ΔFT. No other interaction effects were found.

Table 4.10

Mixed factor measures ANOVA results for scenario specific performance measures for experienced drivers' city scenario

Dependent variable	Gender	Stage	Stage * Gender
ΔNT (^{o}C)	**0.0234**	0.8560	0.9848
ΔFT (^{o}C)	0.6301	0.0917	**0.0016**
ΔNT S.D.	0.3959	**0.0061**	0.5398
MCH	0.2025	< **0.0001**	0.5197
SWAT	0.1134	< **0.0001**	0.8448
Parking duration	**0.0099**	< **0.0001**	0.6972
Parking failure	**0.0384**	< **0.0001**	0.3907

Bolded values denote significant effects

Table 4.11

Descriptive statistics for gender difference for experienced drivers' city scenario (values are in mean (standard deviation))

Dependent variable	Female	Male
ΔNT (^{o}C)	-2.05715 (2.646)	0.64504 (0.958)
Parking duration	56.9712 (27.99)	35.9802 (16.63)
Parking failure	0.64286 (0.655)	0.26190 (0.496)

ΔFT significantly correlated with subjective workload measures strongly, indicating that ΔFT increased when subjective workload increased (Table 4.14). Parking duration also showed a strong correlation with ΔFT. MCH ratings had strong and significant correlations with performance measures. SWAT ratings showed a strong and significant correlation with parking duration.

Table 4.12

Tukey's post hoc results for experienced drivers' city scenario

Dependent variable	Stage	Mean	Group		Dependent variable	Stage	Mean	Group	
ΔNT S.D.	1	0.37	A	B	Parking duration	1	69.40	A	
	2	0.53	A			2	51.47	A	B
	3	0.31		B		3	47.82		B
	4	0.29		B		4	38.73		B
	5	0.30		B		5	36.26		B
	6	0.23		B		6	35.15		B
MCH	1	4.71	A		Parking Failure	1	1.07	A	
	2	3.21		B		2	0.28		B
	3	3.42		B		3	0.28		B
	4	3.00		B		4	0.57	A	B
	5	2.57		B		5	0.28		B
	6	2.35		B		6	0.21		B
SWAT	1	263.3	A						
	2	166.1		B					
	3	191.0		B					
	4	139.5		B					
	5	129.6		B					
	6	120.8		B					

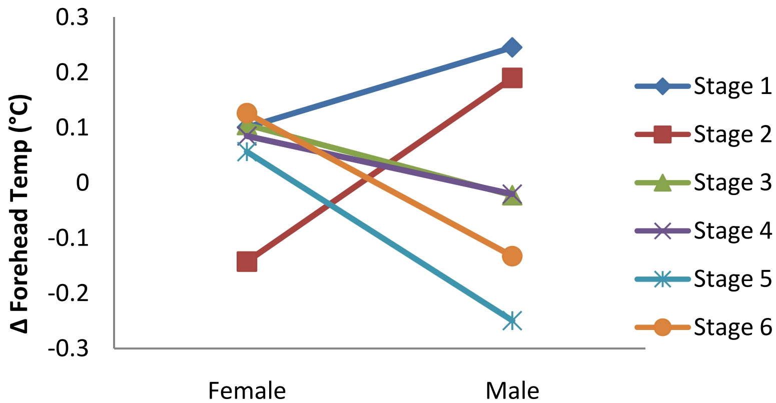


Figure 4.7

Stage by gender interaction effect in ΔFT for experienced drivers' city scenario

Table 4.13

Tukey's post hoc results for the gender by stage interaction effect for experienced drivers' city scenario

Dependent variables	Interaction	Mean	Group		
ΔFT	Female Stage 1	0.10	A	B	C
	Female Stage 2	-0.14	A	B	C
	Female Stage 3	0.10	A	B	C
	Female Stage 4	0.08	A	B	C
	Female Stage 5	0.08	A	B	C
	Female Stage 6	0.12	A	B	C
	Male Stage 1	0.24	A		C
	Male Stage 2	0.19	A	B	C
	Male Stage 3	-0.02	A	B	C
	Male Stage 4	-0.20		B	
	Male Stage 5	-0.24			
	Male Stage 6	-0.13	A	B	C

Table 4.14

Correlation coefficient for experienced drivers for city scenario

	MCH	ΔFT	ΔNT	SWAT	ΔNT S.D.	Parking duration	Parking Failure
MCH	1						
ΔFT	**0.89**	1					
ΔNT	-0.09	-0.29	1				
SWAT	**0.98**	**0.92**	-0.11	1			
ΔNT S.D.	0.36	0.30	0.69	0.33	1		
Parking duration	**0.96**	**0.93**	-0.02	**0.97**	0.52	1	
Parking Failure	**0.86**	0.70	-0.37	0.80	0.09	0.11	1

Bolded values denote significant effects

Highway scenario

For the highway scenario, gender differences were found only for ΔNT, where ΔNT for females was lower (-1.9564°C) than for males (0.7782°C) (Table 4.15). MCH ratings were significantly affected by stage (Table 4.15). However, no trends were identified in the Tukey's post hoc test for MCH ratings (Table 4.16).While a significant stage by gender interaction was found for ΔFT (Figure 4.8), Tukey's was unable to identify any trends in the pairwise comparisons (Table 4.17).

Strong and significant correlations were found between subjective workload measures and between ΔFT and ΔNT S.D. No other strong correlations were found (Table 4.18).

Table 4.15

Mixed factor ANOVA results for experienced drivers' highway scenario

Dependent variable	Gender	Stage	Stage * Gender
ΔNT (^oC)	**0.0429**	0.5333	0.6886
ΔFT (^oC)	0.5823	0.1009	**0.0041**
ΔNT S.D.	0.9723	0.1145	0.2714
MCH	0.6756	**0.0115**	0.2579
SWAT	0.7293	0.0835	0.2690
Lane deviation	0.4329	0.0653	0.2325

Bolded values denote significant effects

Table 4.16

Tukey's post hoc results for experienced drivers' highway scenario

Dependent variable	Stage	Mean	Group
MCH	1	3.21	A
	2	3.14	A
	3	2.78	A
	4	2.14	A
	5	2.35	A
	6	2.07	A

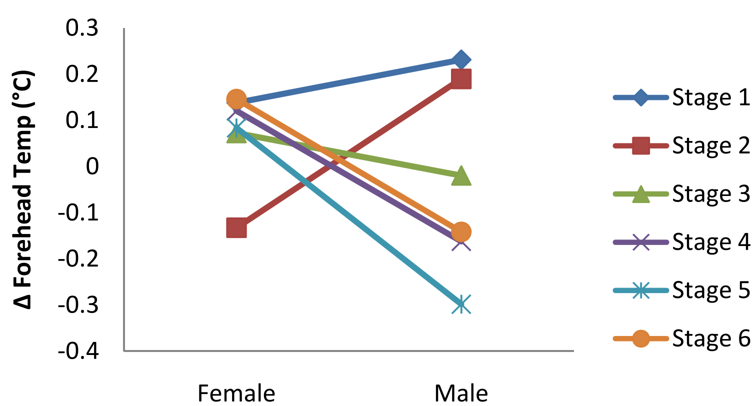


Figure 4.8

Gender by stage interaction effect in ΔFT for experienced drivers' highway scenario

Table 4.17

Tukey's post hoc results for the gender by stage interaction effect for experienced drivers' highway scenario

Dependent variables	Interaction	Mean	Group
	Female Stage 1	0.13	A
	Female Stage 2	-0.13	A
	Female Stage 3	0.07	A
	Female Stage 4	0.1	A
	Female Stage 5	0.08	A
	Female Stage 6	0.14	A
ΔFT	Male Stage 1	0.23	A
	Male Stage 2	0.18	A
	Male Stage 3	-0.01	A
	Male Stage 4	-0.16	A
	Male Stage 5	-0.29	A
	Male Stage 6	-0.14	A

Table 4.18

Correlation coefficients for experienced drivers' highway scenario

	MCH	ΔFT	ΔNT	SWAT	ΔNT S.D.	Lane deviation
MCH	1					
ΔFT	0.69	1				
ΔNT	-0.45	-0.53	1			
SWAT	**0.89**	0.64	-0.56	1		
ΔNT S.D.	0.61	**0.90**	-0.77	0.56	1	
Lane deviation	-0.15	-0.15	-0.16	0.07	0.05	1

Bolded values denote significant effects

Rural scenario

In the rural scenario, gender difference affected ΔNT, with ΔNT for females (-1.6^{o}C) being lower than ΔNT for males (0.8^{o}C) (Table 4.19).

Stage differences were found for the subjective workload measures and drivers' performance (Table 4.19). Subjective workload measures and performance measures were found to be similar across stages 3 to 6, and stage 1 was similar to stage 2 (Table 4.20). For performance measures, stage 1 was similar to stages 3 to 6.

A significant stage by gender interaction effect was found for ΔFT (Table 4.15, Figure 4.9). Male, stage 2 ΔFT (0.2^{o}C) was higher than male, stage 5 ΔFT (-0.3^{o}C). No other interaction effects were found (Table 4.21).

Strong and significant correlations were found between subjective workload measures and performance measures (S-curve and Lane deviation) (Table 4.22). S-curve showed strong and significant correlation with lane deviation. ΔNT showed a strong and significant correlation with ΔNT S.D.

Table 4.19

Mixed factors ANOVA results for experienced drivers' rural scenario

Dependant variable	Gender	Stage	Stage * Gender
ΔNT (^{o}C)	**0.0495**	0.6155	0.6153
ΔFT (^{o}C)	0.4409	0.1701	**0.0048**
ΔNT S.D.	0.7692	0.5363	0.6447
SWAT	0.0611	< **0.0001**	0.4470
MCH	0.1043	**0.0003**	0.1170
S-Curve	1.0000	**0.0134**	0.8251
Lane deviation	0.6427	**0.0018**	0.1983

Bolded values denote significant effects

Table 4.20

Tukey's post hoc results for experienced drivers' rural scenario

Dependent variable	Stage	Mean	Group			Dependent variable	Stage	Mean	Group	
SWAT	1	181.3	A			Lane deviation	1	0.50	A	B
	2	176.2	A	B			2	0.64	A	
	3	120.5	A	B	C		3	0.00		B
	4	116.5		B	C		4	0.28	A	B
	5	102.9			C		5	0.07		B
	6	76.1			C		6	0.07		B
MCH	1	3.28	A	B		S-Curve	1	0.28	A	B
	2	3.71		B			2	0.50	A	
	3	2.57	A	B	C		3	0.14	A	B
	4	2.64	A	B	C		4	0.07		B
	5	2.21	A		C		5	0.07		B
	6	2.00			C		6	0.07		B

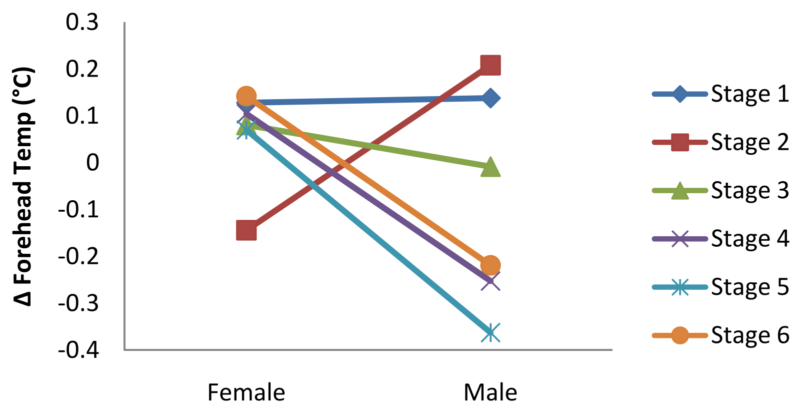


Figure 4.9

Gender by stage interaction effect in ΔFT for experienced drivers' rural scenario

Table 4.21

Tukey's post hoc results for the gender by stage interaction effect for experienced drivers' rural scenario

Dependent variables	Interaction	Mean	Group	
	Female Stage 1	0.12	A	B
	Female Stage 2	-0.14	A	B
	Female Stage 3	0.08	A	B
	Female Stage 4	0.10	A	B
	Female Stage 5	0.07	A	B
	Female Stage 6	0.14	A	B
ΔFT	Male Stage 1	0.13	A	B
	Male Stage 2	0.20		B
	Male Stage 3	-0.008	A	B
	Male Stage 4	-0.25	A	B
	Male Stage 5	-0.36	A	
	Male Stage 6	-0.21	A	B

Table 4.22

Correlation coefficients for experienced drivers' rural scenario

	MCH	ΔFT	ΔNT	SWAT	ΔNT S.D.	S-Curve	Lane deviation
MCH	1						
ΔFT	0.66	1					
ΔNT	-0.18	-0.39	1				
SWAT	**0.96**	0.73	-0.33	1			
ΔNT S.D.	0.19	-0.07	**0.84**	0.08	1		
S-Curve	**0.93**	0.61	0.09	**0.85**	0.50	1	
Lane deviation	**0.92**	0.52	-0.31	**0.86**	0.12	**0.86**	1

Bolded values denote significant effects

4.7.2 NOVICE DRIVER GROUP

In general, ΔNT showed an increasing trend, while ΔFT showed a decreasing trend (Table 4.23). ΔNT S.D. remained unchanged across stages, while movement and total brake time fluctuated. Decreasing trends, in general, were found for collision rates, lost control rates, perception time, and subjective workload measures.

Table 4.23

Descriptive statistics for novice drivers overall session (values are in mean (standard deviation))

Stage	ΔNT (°C)	ΔFT (°C)	ΔNT S.D.	MCH	SWAT (CM)
1	-0.49 (0.71)	0.12 (0.29)	0.24 (0.07)	3.69 (1.46)	189.3 (69.76)
2	-0.11 (1.35)	-0.03 (0.35)	0.25 (0.11)	2.83 (0.94)	134.7 (52.85)
3	0.21 (1.78)	-0.17 (0.47)	0.24 (0.09)	2.50 (0.84)	104.8 (44.15)
4	-0.06 (1.73)	-0.25 (0.54)	0.23 (0.09)	2.19 (1.15)	84.2 (53.92)
5	0.55 (1.91)	-0.06 (0.57)	0.25 (0.13)	2.33 (0.96)	79.3 (46.14)
6	0.52 (1.80)	-0.18 (0.52)	0.20 (0.04)	1.85 (0.75)	69.0 (55.04)
Stage	**Collision (rate of occurrence)**	**Lost control rates (rate of occurrence)**	**Perception time (sec)**	**Movement time (sec)**	**Total brake time (sec)**
1	2.04 (0.91)	0.59 (0.50)	0.94 (0.37)	0.57 (0.47)	1.54 (0.67)
2	1.85 (1.31)	0.35 (0.24)	0.92 (0.46)	0.67 (0.32)	1.56 (0.59)
3	1.40 (1.43)	0.28 (0.31)	0.52 (0.27)	0.45 (0.19)	0.95 (0.42)
4	1.07 (0.75)	0.40 (0.35)	0.76 (0.31)	0.68 (0.32)	1.44 (0.48)
5	0.73 (0.73)	0.04 (0.12)	0.85 (0.29)	0.55 (0.22)	1.37 (0.33)
6	0.78 (0.80)	0.26 (0.35)	0.42 (0.11)	0.38 (0.11)	0.79 (0.19)

No gender differences were found for any of the dependent variables (Table 4.24).

All dependent variables were significantly affected by stage, with the exception of ΔNT S.D. (Table 4.24). ΔNT for stage 1 was lower than stages 5 and 6, while ΔNT was similar across stages 1 to 4 and stages 2 to 6 (Table 4.25). ΔFT for stages 1 and 2 were similar, while stages 3 to 6 were similar. Subjective workload measures in stage 1 were significantly higher than the other stages. Stage 2 was similar with stage 3, while stages 3 to 6 were similar. Lost control rates for stage 1 were lower than stage 5, but were similar to all other stages. Collision rates in stages 5 and 6 were lower than in stages 1 and 2, while stages 3 and 4 were similar to the other stages. Stages 1 and 2 for both perception time and total brake time were similar to stages 4 and 5 and higher than stages 3 and 6 in perception and total brake time.

A significant gender by stage interaction effect was found for ΔFT (Table 4.24, Figure 4.10). Female, stage 1 ΔFT was significantly higher than female, stage 2 ΔFT and male, stage 4 ΔFT. Male, stages 1 and 2 ΔFT were significantly higher than male, stages 3 through 6 ΔFT. No other interaction effects were found.

Thermography (ΔFT and ΔNT) showed strong and significant correlations with subjective workload measures, greater than ±0.83 (Table 4.27). Collision rate was found to be strongly and significantly correlated with MCH ratings (0.89), SWAT ratings (0.93), and ΔNT (-0.88). ΔNT was strongly and significantly correlated with lost control rates (-0.92).

Table 4.24

Mixed factor ANOVA results for novice drivers overall session

Dependant variable	Gender	Stage	Gender * Stage
ΔNT (^{o}C)	0.4678	**0.0103**	0.1270
ΔFT (^{o}C)	0.0799	< **0.0001**	< **0.0001**
ΔNT S.D.	0.3546	0.2641	0.2380
MCH	0.5359	< **0.0001**	0.8631
SWAT	0.6228	< **0.0001**	0.4732
Collision	0.2170	**0.0007**	0.3660
Lost control	0.7515	**0.0020**	0.9820
Perception time	0.4453	< **0.0001**	0.7336
Movement time	0.7266	**0.0345**	0.9024
Total brake time	0.5242	< **0.0001**	0.8450

Bold values indicate significant findings

Table 4.25

Tukey's post hoc results for novice drivers overall session

Dependent variable	Stage	Mean	Group			Dependent variable	Stage	Mean	Group		
ΔNT (^{o}C)	1	-0.49	A			Lost control	1	0.59	A		
	2	-0.11	A	B			2	0.35	A	B	
	3	0.21	A	B			3	0.28	A	B	
	4	-0.06	A	B			4	0.40	A	B	
	5	0.55		B			5	0.04		B	
	6	0.52		B			6	0.26	A	B	
ΔFT (^{o}C)	1	0.12	A			Collision	1	2.04	A		
	2	-0.03	A	B			2	1.85	A		
	3	-0.17		B	C		3	1.40	A	B	
	4	-0.25			C		4	1.07	A	B	
	5	-0.06	A	B	C		5	0.73		B	
	6	-0.18		B	C		6	0.78		B	
SWAT	1	189.3	A			Perception time	1	0.94	A		
	2	134.7		B			2	0.92	A		
	3	104.8		B	C		3	0.52		B	C
	4	84.2			C		4	0.76	A	B	C
	5	79.3			C		5	0.86	A	B	
	6	69.0			C		6	0.42			C
MCH	1	3.69	A			Movement time	1	0.58	A		
	2	2.83		B			2	0.67	A		
	3	2.50		B	C		3	0.45	A		
	4	2.19			C		4	0.68	A		
	5	2.33			C		5	0.54	A		
	6	1.85			C		6	0.38	A		

Table 4.25 Continued

Total brake time	1	1.56	A		
	2	1.56	A		
	3	0.95		B	
	4	1.44	A		
	5	1.37	A		
	6	0.79		B	

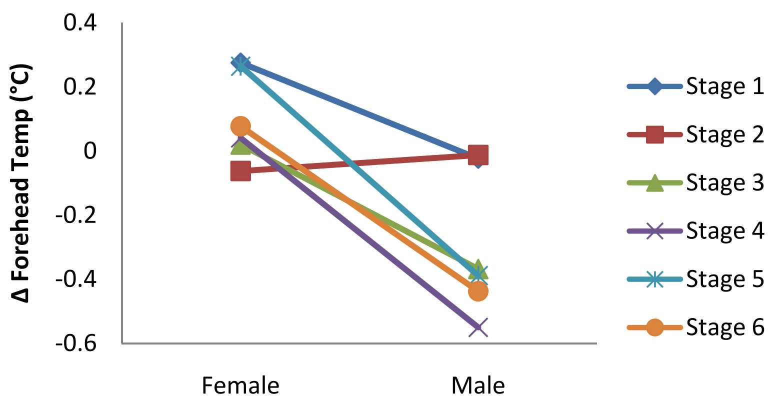


Figure 4.10

Gender by stage interaction effect in ΔFT for novice drivers overall session

Table 4.26

Tukey's post hoc results for the gender by stage interaction effect for novice driver overall session

Dependent variables	Interaction	Mean	Group			
ΔFT (^{o}C)	Female Stage 1	0.27	A	B		
	Female Stage 2	-0.06			C	D
	Female Stage 3	0.01	A	B	C	D
	Female Stage 4	0.04	A	B	C	D
	Female Stage 5	0.26	A	B	C	D
	Female Stage 6	0.07	A	B	C	D
	Male Stage 1	-0.02	A		C	
	Male Stage 2	-0.01	A		C	
	Male Stage 3	-0.36		B		D
	Male Stage 4	-0.55				D
	Male Stage 5	-0.38		B		D
	Male Stage 6	-0.43		B		D

Table 4.27

Correlation coefficients for novice drivers overall session

	MCH	ΔFT	ΔNT	SWAT	ΔNT S.D.	Collision	Lost control	Perception time	Movement time	Total brake time
MCH	1									
ΔFT	**0.88**	1								
ΔNT	**-0.83**	-0.53	1							
SWAT	**0.98**	**0.84**	**-0.87**	1						
ΔNT S.D.	0.50	0.37	-0.36	0.40	1					
Collision	**0.89**	0.67	**-0.88**	**0.93**	0.45	1				
LC	0.68	-0.52	**-0.92**	0.77	-0.01	0.77	1			
PT	0.70	0.68	-0.62	0.63	0.73	0.55	0.31	1		
MT	0.35	0.15	-0.60	0.33	0.64	0.40	0.33	0.80	1	
TBT	0.63	0.53	-0.67	0.57	0.71	0.53	0.38	**0.97**	**0.90**	1

Bolded values denote significant effects

(LC = lost control, PT = perception time, MT = movement time, Total brake time = TBT)

4.7.2.1 INDEPENDENT EVENTS IN EACH SCENARIO

City Scenario

All dependent variables were affected by stage, with the exception of ΔNT S. D (Table 4.28). ΔNT for stage 1 was lower than stages 5 and 6, while parking failure rates for stage 1 were higher than stages 5 and 6 (Table 4.29). Subjective workload measures were found to be similar across stages 3 to 6. MCH ratings for stage 1 were similar to stage 2, while stage 1 SWAT ratings were significantly higher than in the other stages. Parking durations for stage 1 were similar to stage 2, while stages 3 to 6 were similar. Parking failure rates were found to be similar across stages 1 to 4 and across stages 2 to 6, while stage 1 had higher parking failure rates than stages 5 and 6. A significant gender by stage interaction effect was found for ΔFT (Table 2.28, Figure 4.11). Male, stages 1 and 2 was significantly higher than male, stages 3, 4, and 6 (Table 4.30). No gender differences were found.

Thermography (ΔNT and ΔFT) showed strong and significant correlations with subjective workload measures, greater than ±0.81 (Table 4.31). Performance measures (parting duration and failure rate) showed strong and significant correlations with ΔNT and subjective workload measures, greater than ±0.89.

Table 4.28

Mixed factors ANOVA results for novice drivers' city scenario

Dependant variable	Gender	Stage	Gender * Stage
ΔNT (^oC)	0.4112	**0.0078**	0.0904
ΔFT (^oC)	0.0959	< **0.0001**	**0.0003**
ΔNT S.D.	0.2933	0.3266	0.4700
MCH	0.2577	< **0.0001**	0.9350
SWAT	0.4591	< **0.0001**	0.7118
Parking duration	0.2555	< **0.0001**	0.7343
Parking Failure	0.3513	**0.0078**	0.9571

Bolded values denote significant effects

Table 4.29

Tukey's post hoc results for novice drivers' city scenario stages

Dependent variable	Stage	Mean	Group			Dependent variable	Stage	Mean	Group		
	1	-0.79	A				1	4.64	A		
	2	-0.13	A	B			2	3.57	A	B	
	3	0.12	A	B			3	3.00		B	C
ΔNT (^oC)	4	-0.17	A	B		MCH	4	2.85		B	C
	5	0.38		B			5	2.71		B	C
	6	0.27		B			6	2.07			C
	1	0.12	A				1	110.25	A		
	2	-0.03	A	B			2	91.82	A	B	
	3	-0.19		B			3	76.89		B	C
ΔFT (^oC)	4	-0.25		B		Parking duration	4	75.21		B	C
	5	-0.03	A	B			5	69.53		B	C
	6	-0.21		B			6	58.89			C
	1	244.2	A				1	1.71	A		
	2	182.1		B			2	1.42	A	B	
	3	129.1		B	C		3	1.14	A	B	
SWAT	4	108.7			C	Parking failure	4	1.07	A	B	
	5	108.5			C		5	0.92		B	
	6	93.5			C		6	0.78		B	

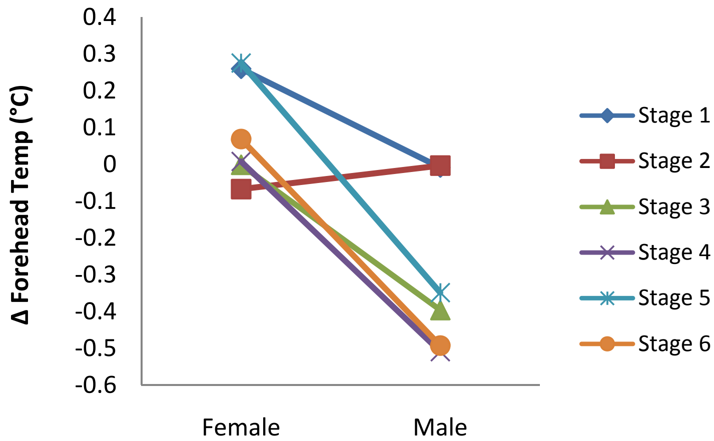


Figure 4.11

Gender by stage interaction effect for ΔFT novice drivers' city scenario stages

Table 4.30

Tukey's post hoc results for the gender by stage interaction effects for novice drivers' city scenario

Dependent variable	Interaction	Mean	Group	
	Female Stage 1	0.25	A	B
	Female Stage 2	-0.06	A	B
	Female Stage 3	-0.0008	A	B
	Female Stage 4	0.0008	A	B
	Female Stage 5	0.27	A	B
	Female Stage 6	0.06	A	B
ΔFT (^{o}C)	Male Stage 1	-0.008	A	
	Male Stage 2	-0.003	A	
	Male Stage 3	-0.39		B
	Male Stage 4	-0.50		B
	Male Stage 5	-0.34	A	B
	Male Stage 6	-0.49		B

Table 4.31

Correlation coefficients for novice drivers' city scenario

	MCH	ΔFT	ΔNT	SWAT	ΔNT S.D.	Parking Duration	Parking Failure Rate
MCH	1						
ΔFT	**0.81**	1					
ΔNT	**-0.89**	-0.58	1				
SWAT	**0.97**	**0.84**	**-0.87**	1			
ΔNT S.D.	0.38	0.51	0.04	0.29	1		
Parking Duration	**0.99**	0.78	**-0.90**	**0.97**	0.34	1	
Parking Failure Rate	**0.98**	0.75	**-0.89**	**0.97**	0.33	**0.99**	1

Bolded values denote significant effects

Highway scenario

For the highway scenario, all dependent variables were affected by stage, with the exception of ΔNT S. D (Table 4.32). ΔNT for stage 1 was lower than stage 5, but was similar across the other stages (Table 4.33). ΔFT for stage 1 was similar to stages 2 and 5, and stages 2 to 5 were similar. MCH ratings were similar for stages 1 to 3 and for stages 2 to 6. SWAT ratings for stage 1 were similar to stage 2, and stages 2 to 6 were similar. Lane deviations for stages 1 and 2 were similar and higher than stage 3. However, lane deviations for stages 1 and 2 were similar to stages 4 to 6. No gender differences were found for any of the dependent variables, and only ΔFT was affected by a gender by stage interaction (Table 4.32, Figure 4.12). Female, stage 1 ΔFT was higher than male, stage 4 ΔFT (Table 4.34). Male, stages 1 and 2 were similar, but were significantly higher than male, stages 4 though 6. No other interaction effects were found for ΔFT.

Significant and strong correlations were found between ΔFT and subjective workload measures, the MCH (0.9) and SWAT (0.87) (Table 4.35). Lane deviation showed strong and significant correlations with ΔFT (0.83) and SWAT (0.9).

Table 4.32

Mixed factor ANOVA results for novice drivers' highway scenario

Dependant variable	Gender	Stage	Gender * Stage
ΔNT (^oC)	0.4306	**0.0201**	0.2719
ΔFT (^oC)	0.0624	< **0.0001**	**0.0002**
ΔNT S.D.	0.4177	0.6751	0.8349
MCH	1.0000	**0.0001**	0.2510
SWAT	0.7969	**0.0003**	0.4583
Lane deviation	0.9408	**0.0016**	0.3207

Bolded values indicate significant results

Table 4.33

Tukey's post hoc results for novice drivers' highway scenario stages

Dependent variable	Stage	Mean	Group		Dependent variable	Stage	Mean	Group	
ΔNT (^oC)	1	-0.47	A		SWAT	1	155.2	A	
	2	-0.11	A	B		2	118.7	A	B
	3	0.26	A	B		3	78.2		B
	4	-0.07	A	B		4	59.7		B
	5	0.69		B		5	62.0		B
	6	0.65	A	B		6	62.9		B
ΔFT (^oC)	1	0.12	A		Lane deviation	1	3.28	A	
	2	-0.04	A	B		2	3.07	A	
	3	-0.16		B		3	1.50		B
	4	-0.26		B		4	1.85	A	B
	5	-0.07	A	B		5	2.00	A	B
	6	-0.16		B		6	1.85	A	B
MCH	1	3.21	A						
	2	2.50	A	B					
	3	2.28	A	B					
	4	1.78		B					
	5	2.07		B					
	6	1.71		B					

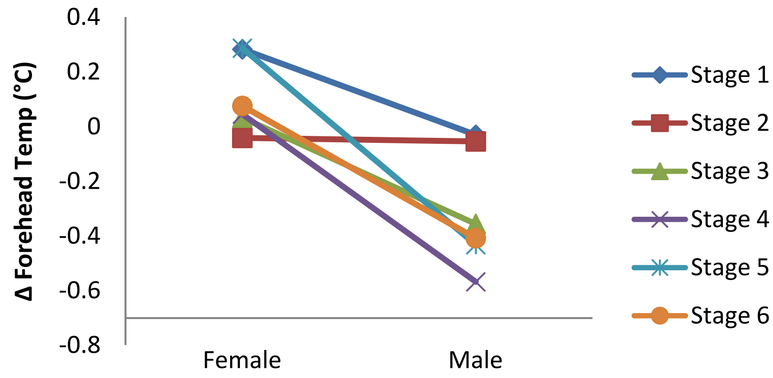


Figure 4.12

Gender by stage interaction effect in ΔFT for novice drivers' highway scenario stages

Table 4.34

Tukey's post hoc results for the gender by stage interaction effect for novice drivers' highway scenario

.Dependent variable	Interaction	Mean	Group	
	Female Stage 1	0.28	A	B
	Female Stage 2	-0.04	A	B
	Female Stage 3	0.03	A	B
	Female Stage 4	0.04	A	B
	Female Stage 5	0.28	A	B
	Female Stage 6	0.07	A	B
ΔFT (^{o}C)	Male Stage 1	-0.03	A	
	Male Stage 2	-0.05	A	
	Male Stage 3	-0.35	A	B
	Male Stage 4	-0.56		B
	Male Stage 5	-0.43		B
	Male Stage 6	-0.40		B

Table 4.35

Correlation coefficients for novice drivers' highway scenario

	MCH	ΔFT	ΔNT	SWAT	ΔNT S.D.	Lane deviation
MCH	1					
ΔFT	**0.90**	1				
ΔNT	-0.72	-0.46	1			
SWAT	**0.95**	**0.87**	-0.79	1		
ΔNT S.D.	0.31	0.28	-0.34	0.28	1	
Lane deviation	0.79	**0.83**	-0.70	**0.90**	0.48	1

Bolded values denote significant effects

Rural scenario

Stage difference affected subjective workload measures, ΔFT, and S-curve performance (Table 4.36). SWAT ratings for stage 1 were significantly higher than in the other stages (Table 4.37). MCH ratings for stage 1 were similar with stage 2 and stages 2 to 6 had similar ratings. S-curve performance for stage 6 was similar to stages 1 to 4. ΔFT was similar across stages 3 to 6, and stage 1 was similar to stages 2 and 5. No gender differences were found for any of the dependent variables. A gender by stage interaction effect was found for ΔFT (Table 3.6, Figure 4.13). Female, stages 1 and 5 ΔFT were higher than male, stage 4 ΔFT (Table 4.38). Male stage 1 ΔFT) was higher than male, stage 4 and 6 ΔFT. Male, stage 2 ΔFT was higher than male, stage 3 to 6 ΔFT.

A strong and significant correlation was found between ΔNT and SWAT (-0.83) (Table 4.39). S-curve performance showed strong and significant correlations with ΔNT (-0.89) and SWAT (0.82). MCH correlated strongly and significantly with ΔFT (0.92), while ΔNT S.D. correlated significantly with lane deviation (0.88).

Table 4.36

Mixed factors ANOVA results for novice drivers' rural scenario

Dependant variable	Gender	Stage	Gender * Stage
ΔNT (^oC)	0.5768	0.0697	0.1325
ΔFT (^oC)	0.0851	**0.0001**	**< 0.0001**
ΔNT S.D.	0.5933	0.7915	0.1980
MCH	0.9570	**0.0004**	0.6278
SWAT	0.9134	**< 0.0001**	0.8144
Lane deviation	0.7285	0.4653	0.9796
S-Curve	0.8142	**0.0149**	0.9662

Bolded values denote significant effects

Table 4.37

Tukey's post hoc results for novice drivers' rural scenario stages

Dependent variable	Stage	Mean	Group			Dependent variable	Stage	Mean	Group	
ΔFT (^{o}C)	1	0.12	A			MCH	1	3.21	A	
	2	-0.02	A	B			2	2.42	A	B
	3	-0.16		B	C		3	2.21		B
	4	-0.25			C		4	1.92		B
	5	-0.07	A	B	C		5	2.21		B
	6	-0.16		B	C		6	1.78		B
SWAT	1	168.6	A			S- Curve	1	0.64	A	
	2	103.2		B	C		2	0.42	A	B
	3	107.0		B			3	0.35	A	B
	4	84.3		B	C		4	0.42	A	B
	5	67.5		B	C		5	0.00		B
	6	50.5			C		6	0.21	A	B

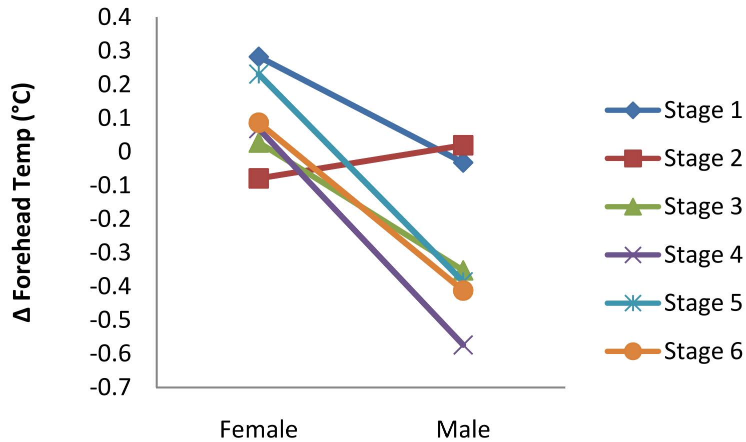


Figure 4.13

Gender by stage interaction effect in ΔFT for novice drivers' rural scenario

Table 4.38

Tukey's post hoc results for the gender by stage interaction effect for novice drivers' rural scenario

Dependent variable	Interaction	Mean	Group		
	Female Stage 1	0.28	A	B	
	Female Stage 2	-0.07	A	B	C
	Female Stage 3	0.02	A	B	C
	Female Stage 4	0.06	A	B	C
	Female Stage 5	0.23	A	B	
	Female Stage 6	0.08	A	B	C
ΔFT (^oC)	Male Stage 1	-0.03	A	B	
	Male Stage 2	0.01	A		
	Male Stage 3	-0.35		B	C
	Male Stage 4	-0.57			C
	Male Stage 5	-0.38		B	C
	Male Stage 6	-0.41			C

Table 4.39

Correlation coefficients for novice drivers' rural scenario

	MCH	ΔFT	ΔNT	SWAT	ΔNT S.D.	S-curve performance	Lane deviation
MCH	1						
ΔFT	**0.92**	1					
ΔNT	-0.70	-0.49	1				
SWAT	**0.93**	0.74	**-0.83**	1			
ΔNT S.D.	0.08	-0.25	-0.56	0.43	1		
S-curve performance	0.63	0.41	**-0.89**	**0.82**	0.71	1	
Lane deviation	0.32	0.01	-0.72	0.59	**0.88**	**0.84**	1

Bolded values denote significant effects

4.7.3 SKILL MASTERY AND LEARNING

Skill mastery for experienced drivers was assessed by comparing stages for specific skills. To measure novice drivers learning and skill master, their performance in each stage of each scenario was compared to the poorest performance level of the experienced drivers. This served as a conservative value for minimum performance needed to master the driving skill.

Perception time

No difference in perception time between the stages was found for experienced drivers (Table 4.6, p-value = 0.0823), indicating experienced drivers adapted to the new driving environment during stage 1 (Figure 4.14). Perception time performance for experienced drivers was poorest for stage 2 (0.33 seconds). Novice drivers, however, did differ in perception time across the stages (Table 4.24, p-value < 0.0001). Perception times for stages 3 and 6 were the lowest (0.52 and 0.42 seconds respectively), and differed significantly from stages 1 and 2 (Table 4.23). When comparing experienced drivers "worst" performance to novices, only stages 3 and 6 showed similar performance (not statistically different), though expert performance was slightly lower than that of novices for perception time (Table 4.40). Trends for the two driver groups were not similar (Figure 4.11).

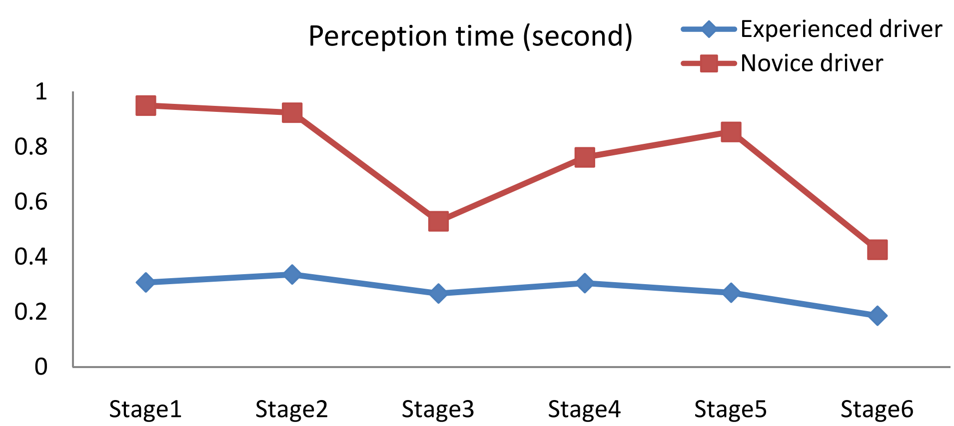


Figure 4.14

Trends of perception time for experienced and novice drivers

Table 4.40

Comparisons between experienced and novice drivers for perception time

Experienced drivers' stage	Novice drivers' stage	p-values
2	1	**< 0.0001**
2	2	**< 0.0001**
2	3	0.7032
2	4	**0.0017**
2	5	**< 0.0001**
2	6	0.9988

Bolded values denote significant differences

Movement time

As with perception time, no differences between stages in movement time were found for experienced drivers (Table 4.6, p-value = 0.1489), though differences were found for novice drivers (Table 4.24, p-value = 0.0345). Again, stage 2 was associated with the poorest performance level (0.52 seconds) for expert drivers. Tukey's post hoc tests were unable to identify differences in the pairwise comparisons for the stages for novice drivers (Table 4.25). Trends in this performance metric for both groups were similar, with an increase in movement time at stage 2 for both groups and at stage 4 for novice drivers (Figure 4.15). When comparing expert drivers "worst" performance with the novices best performance, no differences were found between the two groups (Table 4.41), though there was a trend for experts to have lower values (better performance) than novices.

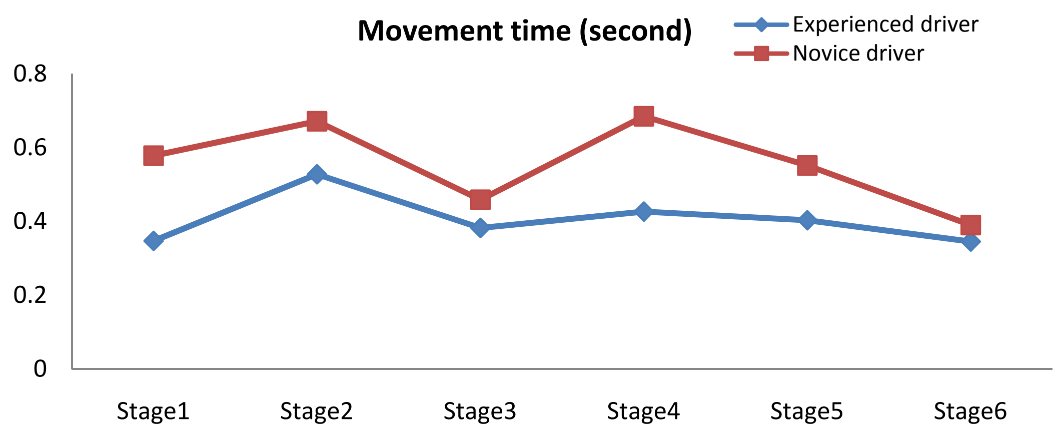


Figure 4.15

Trends of movement time for experienced and novice drivers

Table 4.41

Comparisons between experienced and novice drivers for movement time

Experienced drivers' stage	Novice drivers' stage	p-values
2	1	1.0000
2	2	0.9464
2	3	0.9999
2	4	0.9034
2	5	1.0000
2	6	0.9590

Bolded values denote significant differences

Total brake time

Total brake time was found to be affected by stage for both experienced (Table 4.6, p-value = 0.0101) and novice drivers (Table 4.24, p-value = < 0.0001). Tukey pairwise comparisons found that for experienced drivers, stage 2 differed from stage 6, with stage 2 having the longest total braking time and stage 6 having the shortest total braking time (Figure 4.16). Stage 1 was similar with all other stages, indicating that there is no difference between stages from the initial exposure (stage 1). Again, performance was poorest in stage 2 for experienced drivers. For novice drivers, stages 3 and 6 had significantly shorter braking times than the other stages, and this trend was similar to what was found for perception time. As was found with perception time, novice performance during stages 3 and 6 was similar to that of experienced drivers' poorest performance (Table 4.42).

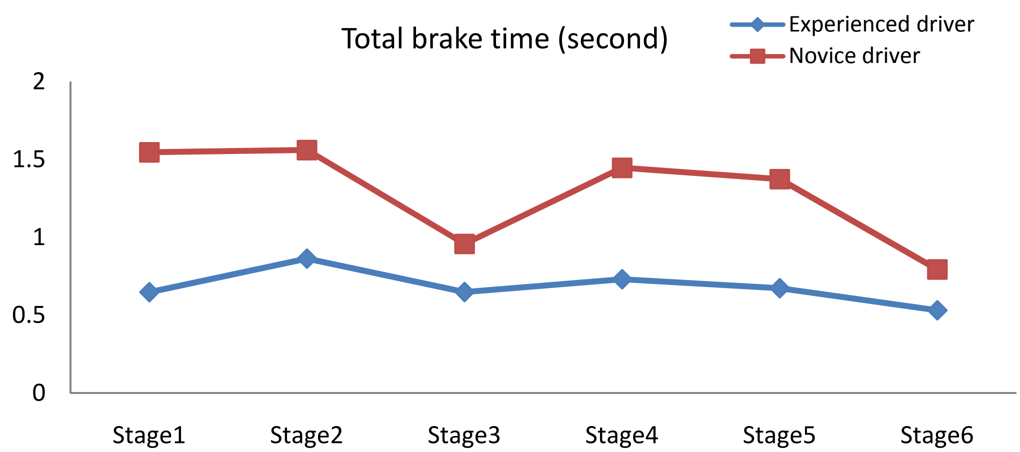


Figure 4.16

Trends of total brake time for experienced and novice drivers

Table 4.42

Comparisons between experienced and novice drivers for total brake time

Experienced drivers' stage	Novice drivers' stage	p-values
2	1	**0.0006**
2	2	**0.0004**
2	3	1.000
2	4	**0.0069**
2	5	**0.0337**
2	6	1.000

Bolded values denote significant differences

Collision rate

Stage effects were found for both experienced (Table 4.6, p-value = 0.0136) and novice drivers (Table 4.24, p-value = 0.0007) with respect to collision rates. For experienced drivers, tukey's post hoc test identified that stage 1 was significantly higher than stage 6, while stages 2 to 6 were similar (Table 4.7, Figure 4.17). Given that stage 1 was associated with skill mastery, stage 2 was selected as the poorest performance level. For novice drivers, a decreasing trend in collision rate was observed (Figure 4.17). Post hoc tests found that stages 1 and 2 were significantly higher than stages 5 and 6, and stages 3 to 6 were similar (Table 4.25). Only stages 1 and 2 of novice drivers differed from experienced drivers stage 2 performance level (Table 4.43).

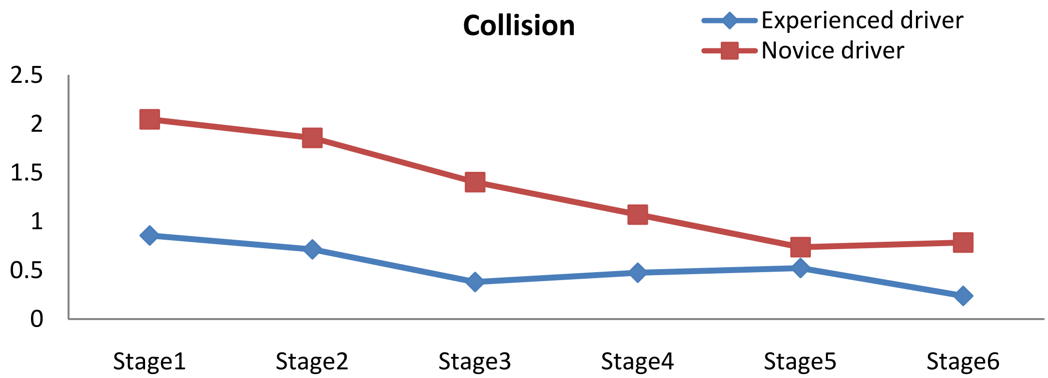


Figure 4.17

Trends of collision for experienced and novice drivers

Table 4.43

Comparisons between experienced and novice drivers for collision rate

Experienced drivers' stage	Novice drivers' stage	p-values
2	1	**0.0017**
2	2	**0.0150**
2	3	0.5095
2	4	0.9901
2	5	1.0000
2	6	1.0000

Bolded values denote significant differences

Lost control rates

Stage effects were found for both experienced (Table 4.6, p-value = 0.0229) and novice drivers (Table 4.24, p-value = 0.0020) with respect to lost control rates. Again, skill mastery was occurring in stage 1 for experienced drivers (Table 4.7). Stage 2 was significantly higher than stage 5, but both stages 2 and 5 were similar to stage 1. As stage 1 was their first exposure to this skill, it is reasonable to conclude that the experienced drivers were unable to master this skill. Again, stage 2 was associated with the poorest performance for experienced drivers. Novice drivers' best performance was found in stage 5 (0.04), which was similar to the other stages, with the exception of stage 1 (Table 4.25). The trend in lost control rates for both experienced and novice drivers were decreasing, though there was fluctuation in these rates (Figure 4.18). However, novice drivers' performance was closer to experienced drivers' performance after stage 1. Experienced drivers stage 2 was different from novice stage 1 but was similar to the other novice stages (Table 4.44).

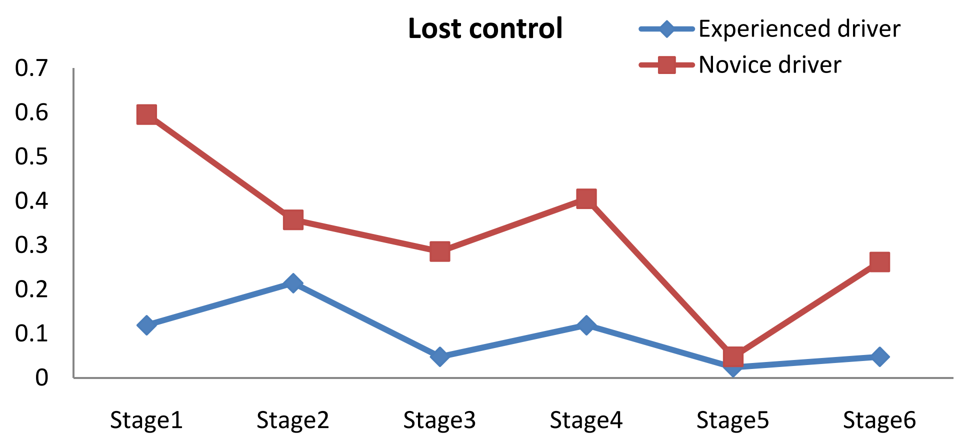


Figure 4.18

Trends of lost control for experienced and novice drivers

Table 4.44

Comparisons between experienced and novice drivers for lost control rates

Experienced drivers' stage	Novice drivers' stage	p-values
2	1	**0.0176**
2	2	0.9635
2	3	0.9999
2	4	0.7868
2	5	0.8984
2	6	1.0000

Bolded values denote significant differences

Parking duration (city scenario)

Parking duration was found to be affected by stage for both experienced (Table 4.10, p-value < 0.0001) and novice drivers (Table 4.28,p-value < 0.0001). Tukey pairwise comparisons found that for experienced drivers, stage 1 differed from stages 3 through 6 and stages 2 through 6 were similar (Table 4.12). Because stage 1 was associated with skill mastery, stage 2 was selected as the poorest performance level. A decreasing trend in parking duration was found for both driving groups, particularly across stages 1 to 3 (Figure 4.19). For novice drivers, stages 1 and 2 were similar, but differed from stages 3 through 6 (Table 4.29). Post hoc tests found that stage 2 of experienced drivers was similar with novice drivers' stages 5 and 6 performance level (Table 4.45).

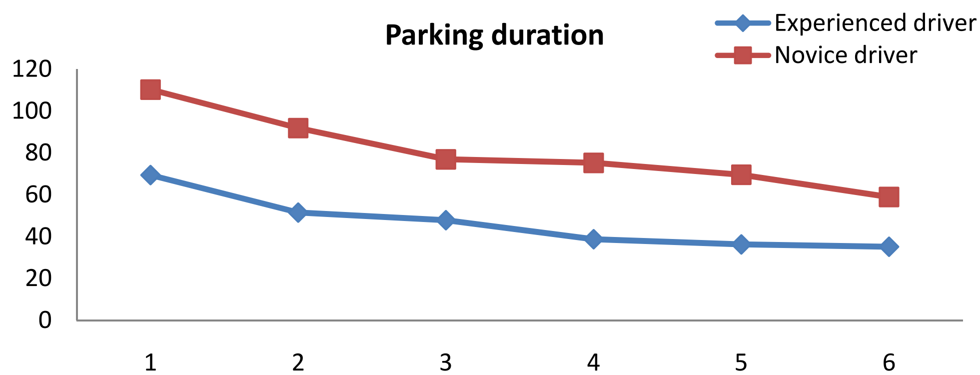


Figure 4.19

Trends of parking duration for experienced and novice drivers

Table 4.45

Comparisons between experienced and novice drivers for parking duration

Experienced drivers' stage	Novice drivers' stage	p-values
2	1	**< 0.0001**
2	2	**0.0002**
2	3	**0.0168**
2	4	**0.0252**
2	5	0.0861
2	6	0.4768

Bolded values denote significant differences

Parking Failure Rate (City Scenario)

Stage effects were found for both experienced (Table 4.10, p-value < 0.0001) and novice drivers (Table 4.28, p-value = 0.0078) with respect to parking failure rate. For experienced drivers, tukey post hoc tests revealed that stage 1 was higher than the other stages, while stages 2 through 6 were similar (Table 4.12). Stage 4 was the worst performance stage, following skill mastery in stage 1. Experienced drivers' maintained a consistent performance level, while novices exhibited a decreasing trend (Figure 4.20). For novice drivers, post hoc tests found that stage 1 was significantly higher than stages 5 and 6, while stages 2 to 6 were similar (Table 4.29). Stages 1 and 2 of novice drivers differed from experienced drivers' stage 4 (Table 4.46).

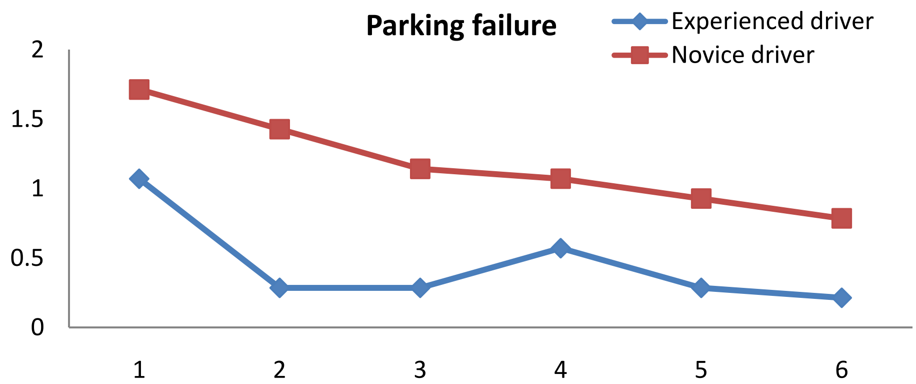


Figure 4.20

Trends of parking failure for experienced and novice drivers (average number of occurrences)

Table 4.46

Comparisons between experienced and novice drivers for parking failure rate

Experienced drivers' stage	Novice drivers' stage	p-values
4	1	**0.0011**
4	2	**0.0470**
4	3	0.5231
4	4	0.7170
4	5	0.9607
4	6	0.9995

Bolded values denote significant differences

Lane deviation (Highway Scenario)

No difference was found between stages for experienced drivers (Table 4.15, p-value = 0.0653), while novice drivers were affected by stage (Table 4.32, p-value = 0.0016). For experienced drivers, performance was consistent across stages, with the poorest performance occurring at stage 4 (Figure 4.21). For novice drivers, stages 1 and 2 were similar and stages 4 through 6 were similar (Table 4.33). Lane deviation decreased from stages 1 to 3, then increased until stage 5 (Figure 4.21). Novice drivers did not improve their performance for highway lane deviation according to a post hoc comparison (Table 4.33) that showed stage 1 was similar with the other stages, with the exception of stage 3 only. Experienced drivers stage 4 was similar to novice drivers stage 3, 4, and 6 (Table 4.47).

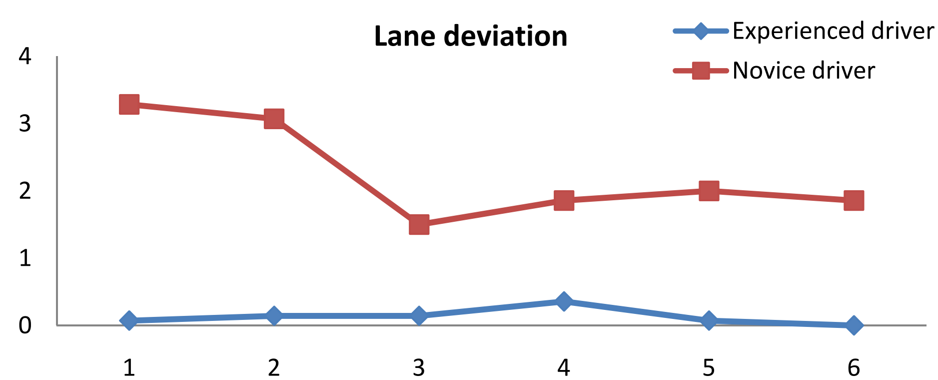


Figure 4.21

Trends of lane deviation for experienced and novice drivers for lane deviation for highway scenario (average number of occurrences)

Table 4.47

Comparisons between experienced and novice drivers for lane deviation (highway scenario)

Experienced drivers' stage	Novice drivers' stage	p-values
4	1	**< 0.0001**
4	2	**< 0.0001**
4	3	0.3426
4	4	0.0560
4	5	**0.0221**
4	6	0.0560

Bolded values denote significant differences

Lane deviation (Rural Scenario)

Stage effects were found experienced drivers (Table 4.19, p-value = 0.0018). For experienced drivers, stage 1 was, however, similar to the other stages according a post hoc comparison, indicating that their performance level was not changed from initial stage although stages effect was significant (Table 4.20). Novice drivers, however, did not differ in lane deviation across the stages (Table 4.36, p-value = 0.4653). Stage 2 was associated with the poorest performance level for expert drivers (Table 4.20, Figure 4.22). When this study compared experienced drivers' poorest performance (stage 2) with the novices' performance across the stages, experienced drivers' stage 2 was similar to all novice drivers stages (Table 4.48).

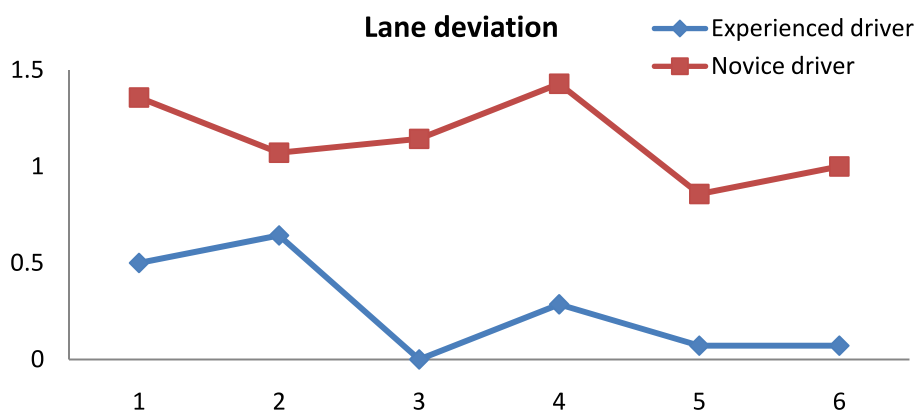


Figure 4.22

Trends of lane deviation for experienced and novice drivers in rural scenario (average number of occurrences)

Table 4.48

Comparisons between experienced and novice drivers for lane deviation in rural scenario

Experienced drivers' stage	Novice drivers' stage	p-values
2	1	0.2934
2	2	0.9187
2	3	0.8026
2	4	0.1715
2	5	0.9997
2	6	0.9772

S - Curve (Rural Scenario)

Stage effects were found for both experienced (Table 4.19, p-value = 0.0134) and novice drivers (Table 4.36, p-value = 0.0149) with respect to S-curve failure rate. For experienced drivers, stage 2 was significantly higher than stages 4 to 5, but stage 1 was similar to all stages including stage 2. For novice drivers, stage 1 was significantly higher than stage 5, but stages 2 to 6 were similar. Experienced drivers' stage 2 was associated with the poorest performance level (Figure 4.23). Post hoc test found there were no difference between experienced drivers' stage 2 and novice drivers' stages 1 to 6 (Table 4.49).

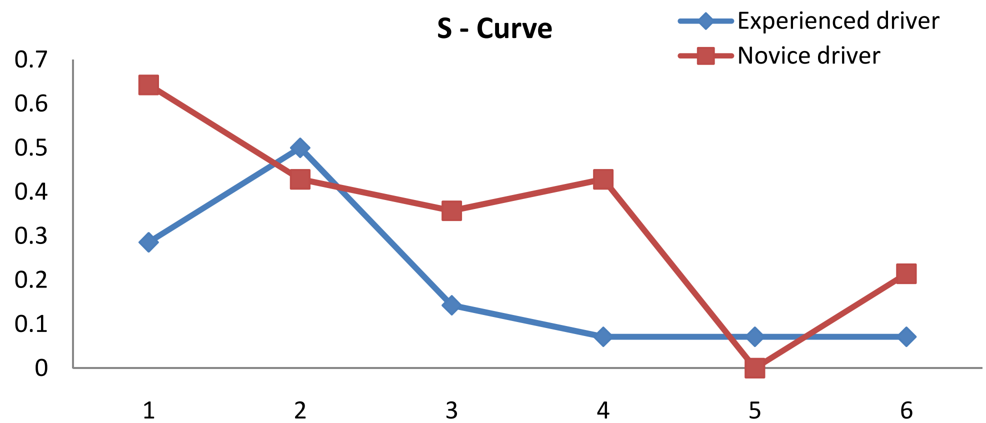


Figure 4.23

Trends of lane deviation for experienced and novice drivers (average failed number)

Table 4.49

Comparisons between experienced and novice drivers for S-curve performance

Experienced drivers' stage	Novice drivers' stage	p-values
2	1	0.9991
2	2	1.0000
2	3	0.9991
2	4	1.0000
2	5	0.0944
2	6	0.8251

4.7.4 THE TREND OF MENTAL WORKLOAD MEASURES

Subjective workload measures decreased for both experienced and novice drivers (Figure 4.24, Figure 4.25), as supported by slope computations (Table 4.50). Only, MCH ratings for novice drivers increased slightly in stage 5 (Figure 4.25). ΔNT showed an increasing trend with a fluctuation (Table 4.50). Particularly, stage 4 was lower than the other stages except stage 1 (Figure 4.26).

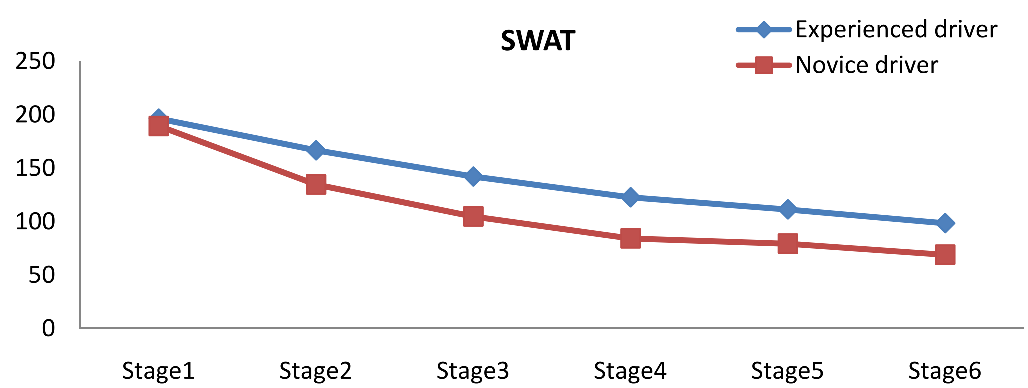


Figure 4.24

SWAT rating trends for experienced and novice drivers

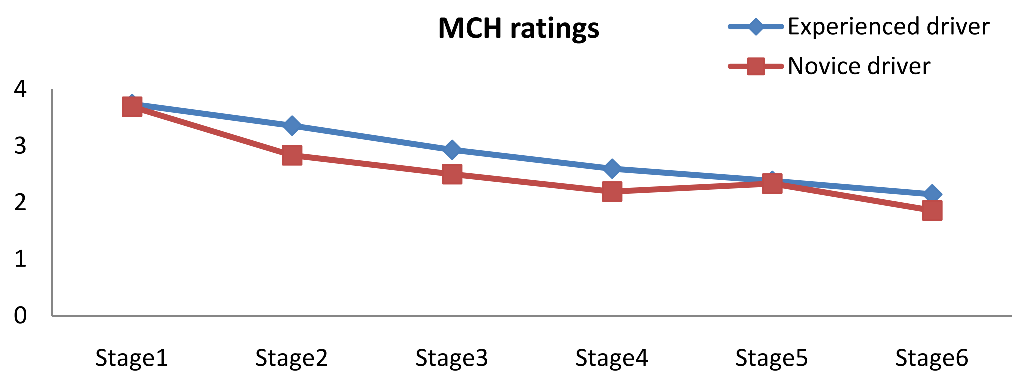


Figure 4.25

MCH rating trends for experienced and novice drivers

Table 4.50

Comparison between experienced and novice drivers for subjective mental workload measures slopes and thermography

Mental workload measure	P – value	Experienced (Mean (S.D.))	Novice (Mean (S.D.))
MCH	0.9323	-0.32 (0.25)	-0.31 (0.21)
SWAT	0.5520	-19.25 (15.6)	-22.52 (12.9)
ΔNT	0.2269	0.03 (0.37)	0.19 (0.31)

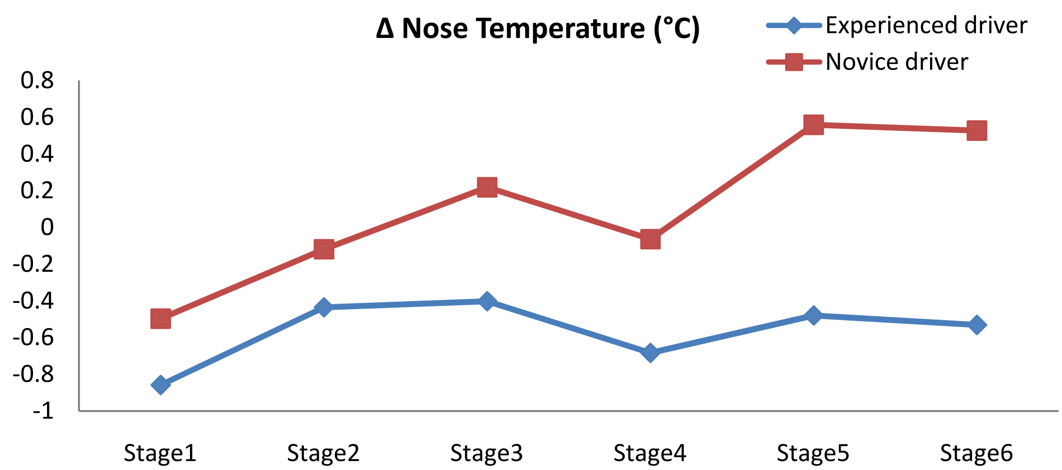


Figure 4.26

Trend of the ΔNT for experienced and novice drivers for overall stages

4.8 DISCUSSION

This study employed a driving simulator to provide a complex training environment to investigate the relationship between facial temperature changes and human mental workload during learning. Experience level (novice and expert drivers) was also considered in quantifying the relationship between thermal images and mental workload during complex task completion.

In general, thermography and subjective workload measures indicated that experienced and novice drivers had similar workload levels (Table 4.1), though performance measures for experienced drivers' was better than novices; notable exceptions included over-speed, average speed, and blinker use where performance was comparable across the experience groups.

This study selected performance measures to indicate differences between experience and novice drivers, to investigate novice drivers' learning and skill mastery. There are no evaluation tools to indicate whether novice drivers learn a driving skill. For this study, experienced drivers' performance data was used as criteria to indicate whether novice drivers mastered the skills. When humans experience a new environment, adaptation occurs. For example, human eyes will adjust to the lighting levels when entering a room (Sanders and McCormick, 1993).Therefore, the stage following adaptation, or skill mastery, with the poorest performance level for experienced drivers was selected as the criterion for assessing novice learning. Further discussion of the results and the specific hypotheses are presented below.

Hypothesis1: Novice and experienced drivers will perform differently.

In general, this hypothesis was supported. As mentioned previously, performance for experienced drivers exceeded that of novice drivers with the exception of over-speed, average speed, and blinker use, indicating experienced drivers were driving in safer manner (Dorn and Barker, 2005). Over-speed and average speed were affected by internal simulator warnings, (e.g., "please speed up" or "warning over speed"). Therefore, differences in these performance measures may have been overshadowed by this confounding event. Blinker use was also affected by scenario design as a finite number of instances required the use of the blinker. Blinker use during events outside of those scripted into the scenario cannot be estimated, nor can any performance differences between experience groups for these events. Therefore, the lack of differences in these performance measures may have been due to experimental design rather than being a true reflection of how these two groups differs for these dimensions. The scenarios created were intended to be representative of real-world driving scenarios, however, they are simulations and these limitations cannot be overcome regardless of the experimental design.

Hypothesis 2: Novice and experienced drivers will improve their performance until they have mastered the skill in the simulation environment. After this point, their performance will improve slightly or remain constant.

Experienced drivers

All performance metrics indicated that experience drivers adapted the simulated environment within the first stage. Experienced drivers needed no adaptation period for several skills (braking performance (perception, movement, and total braking time), lost control rate, S-curve, and land deviation (highway and rural scenario)). Other skills required the first stage for skill mastery in the simulator (collision rate, parking duration, and parking failure rate).

For those skills that required no adaptation period, error rate increased slightly from stage 1 to 2 although there were no significant differences between stages, indicating that experienced drivers' negligence increased the error rate at skill-based performance levels (Rasmussen, 1983, 1986, 1993). However, experienced drivers exhibited decreasing error rates when they needed an adaptation period, because drivers were aware of their poorer performance.

Novice drivers

In general, novice drivers' performance remained constant across stagesafter they acquired the skills by training/learning stages. Several performance metrics indicated that novice drivers mastered skills during the experiment, with the exception of the braking performance (perception and total brake time). Both performance measures showed that stages 3 and 6 reached the experienced drivers' performance level, though this performance level was not maintained. This finding indicates that participants mis-manipulated their vehicle due to misinterpretation of a situation at rule-based performance level (Rasmussen, 1983, 1986, 1993). Novice drivers did not realize, or realized later than experienced drivers, emergency situations, as indicated by novice driver's slower perception times.

Total brake time showed a similar trend with perception time with a strong and significant correlation (0.97), while movement time did not fluctuate across stages. This finding indicates that perception time was the driving factor in braking performance, and that to reduce total brake time, one must reduce perception time. Te Velde et al. (2005) indicated that young people required more time to obtain adequate information for decision making because of the lack of experience in visual tracking. Novice drivers were inexperienced and younger people in this study. Braking performance for stage 6 was similar to the poorest performance of experienced drivers, supporting the premise that novice drivers need more training to reach the experience drivers performance level.

Movement time, lane deviation (rural), and S-curve showed that there were no differences between stages, indicating that novice drivers did not need any training or need a training less than 15 minutes.

Stages 1 and 2 were skill mastery stages for collision rates, parking performance (duration and failure rate), and lane deviation (highway scenario), while lost control needed only one stage for skill mastery. These performance measures reached the experienced performance level following skill mastery, indicating novice driver's improved their performance level during the training stages and maintained that performance level. Parking duration and failure rate correlated strongly and significantly (0.99), indicating that longer parking duration occurred when a novice driver had not learned the skill, indicated by the failure rate.

Hypothesis 3: The process of training and practicing to obtain a driving task skill will reduce mental workload over time

Subjective mental workload measures, for both experienced and novice drivers, decreased (Figures 4.21 and 4.22) and thermal readings (ΔNT) generally increased (Figure 2.23), indicating that drivers' mental workload decreased over the course of the test session. SWAT and MCH ratings indicated that experienced drivers experienced higher mental workload in stages 1 and 2. In particular, MCH ratings were affected by a gender by stage interaction effect, with females having higher mental workload ratings initially. A possible explanation for this finding is video game experience. A number of driving video games may have provided males with "experience" in a simulated driving environment, thereby reducing mental workload for the male participants.

Novice drivers had similar mental work load trends as experienced drivers'. SWAT and MCH ratings decreased across stages, and there were no significant differences between experienced and novice drivers ratings (p-value = 0.9323 for MCH ratings, p-value = 0.5520 for SWAT). These findings indicated mirror those in chapter 2 (Table 2.3) and chapter 3 (Table 3.1). As in those chapters, the decrease in the ratings may not be associated with performance or actual workload, but rather in the perception of becoming familiar with the simulator environment.

Thermal readings were not significantly different between stages for the experienced drivers, indicating that mental workload did not change throughout the test session, though an increasing trend was observed (Table 4.5). This trend is consistent with previous findings, though it is likely that additional time in the simulator may be needed to fully quantify nose temperature changes in this environment (Kang et al, 2006; Kang and Babski-Reeves, 2008). Nose temperature readings for novice drivers did increase significantly from stage 1 to 2 indicating a reduction in mental workload consistent with previous findings (Kang et al, 2006; Kang and Babski-Reeves, 2008).

It is interesting to note the decrease in nose temperature associated with stage 4 for both groups. Exposure to driving scenarios was balanced across all participants, eliminating the possibility that a single scenario was responsible for the increased mental workload that occurred at stage 4. Therefore, factors external to the research study played a role in this finding.

As subjective ratings did not coincide with thermal readings (or performance measures as discussed earlier), it is possible that participants "falsely" indicated a reduction in mental workload. While it may be true that participants perceived that the task became easier with repeated exposure, mental workload increased due to fatigue and boredom may not have been perceived. This finding is similar to previous studies (Ross et al, 1975; Kahneman and Tversky, 1973; Andre and Wickens, 1995).

Fatigue has also been found to severely affect driving performance (McDonald, 1984; Storie, 1984; Nilsson et al, 1997; Galinsky et al., 1993; Skipper and Wierwille, 1986), and has been found to peak at 80 minutes (Nilsson et al, 1997). Stage 4 of this study began at the 60 minute mark and ended at minute 75, near this "peak fatigue" point (Figure 4.27).

Nilsson et al. (1997) found that fatigue ratings increased until 80 minutes of driving then fluctuated for the remaining testing time. When comparing the time scale of Nilsson et al.'s study with that of this study, it is clear that 80 minutes occurred following stage 4 (Figure 4.27), which would coincide with the highest level of fatigue experienced in Nilsson et al.'s study. The presence of a high rate of fatigue in stage 4 would have impacted performance, as was seen in the current study's data. For example, Figures 4.28 shows how blinker usage decreased in stage 4 (higher percentage of missed opportunities for misses) for both experienced and novice drivers. Nilsson et al.'s studies also found that fatigue levels for the remaining testing time never reached the maximum value observed at 80 minutes. The current study's data was found to follow a better performance trend for the remaining two stages (stage 5 and 6).

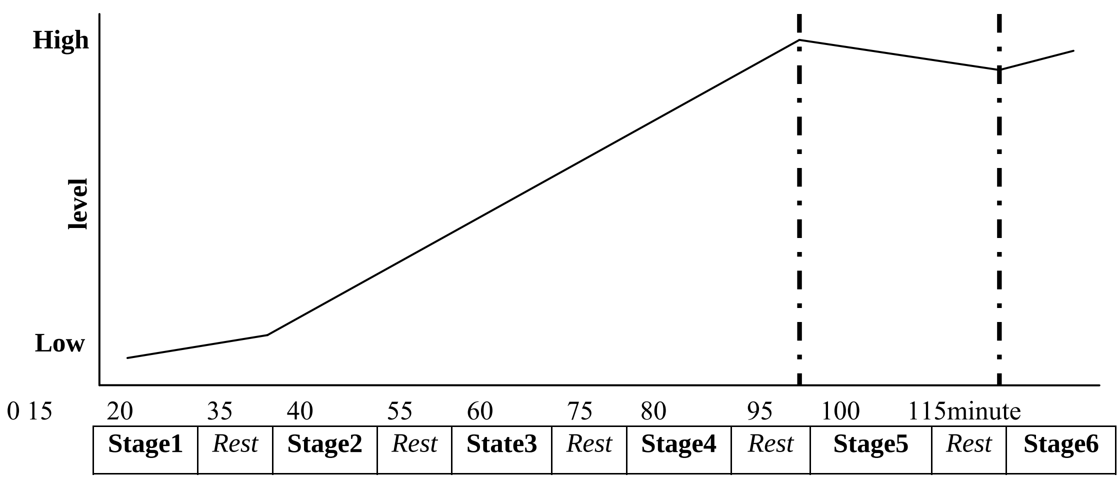
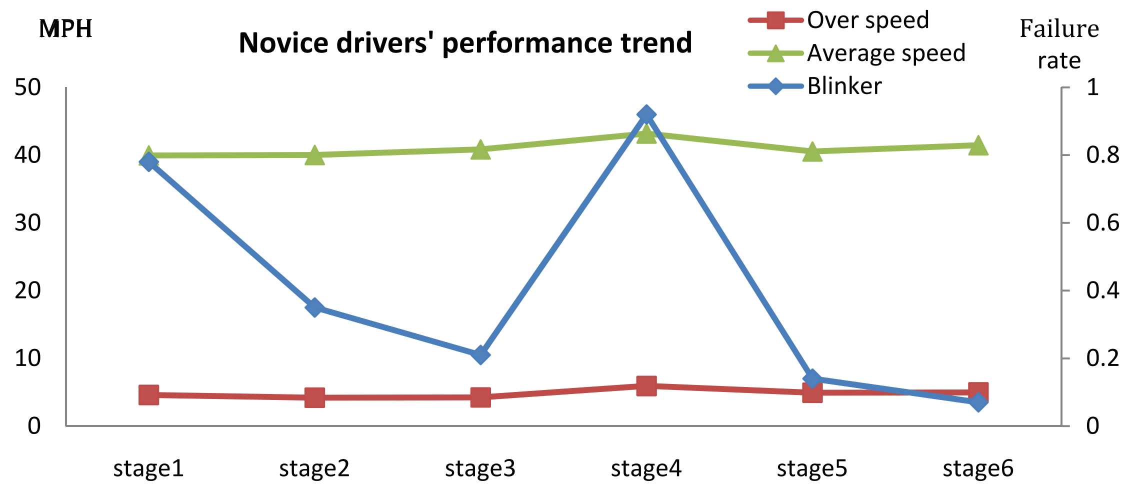


Figure 4.27

Fatigue level fluctuations by time (minutes) (Nilsson, 1997)



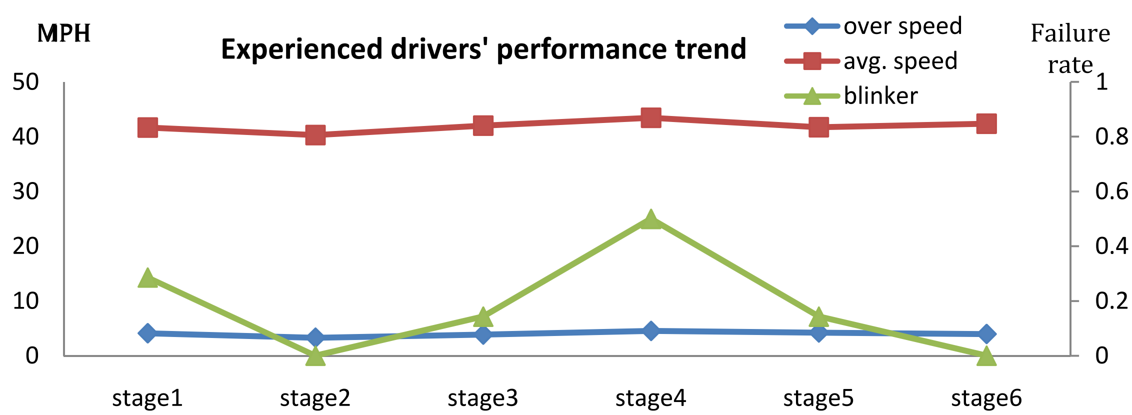


Figure 4.28

Trend of the drivers' performances that were found no differences between experienced and novice drives.

Hypothesis4: Facial temperature readings will have a relationship with a learner's performance and subjective mental workload measures. A learner's performance will have a relationship with mental workload.

In general, support for this hypothesis was found, though findings for forehead temperature are contradictory to previous studies for novice drivers as learners (Kang et al, 2006; Kang and Babski-Reeves, 2008; Veltman and Vos, 2005, Stoll, 1964). Strong and significant correlations were found between performance and ΔNT (greater than -0.88, Table 4.51) and between performance and subjective workload ratings (greater than 0.82, Table 4.51). These findings were supported by previous research, indicating workload decreased, while performance improved (Kang et al, 2006; Kang and Babski-Reeves, 2008). Further, thermal readings (ΔNT and ΔFT) correlated strongly with subjective workload measures (greater than ± 0.83, Table 4.52), again consistent with previous studies (Kang et al, 2006; Kang and Babski-Reeves, 2008). This finding supports the use of thermography as a valid and reliable mental workload measure because research has found subjective workload measures to be valid and reliable workload measures (Roscoe, 1992).

As was found in chapter 2, SWAT ratings were more closely correlated with ΔNT than MCH ratings, as was found in chapter 2 and in previous research (Kang and Babski-Reeves, 2008). However, SWAT ratings were more closely related with only 3 performance measures versus 5 performance measures for the MCH ratings (Table 4.51). ΔNT was also correlated with more performance measures than MCH ratings. Brake performance (perception, movement, total brake time) did not show any strong and significant correlation with subjective workload measures and thermography. This was likely due to sampling rate issues. Braking performance was less than 2 seconds, but thermal data was collected every 10 seconds, at 1 sample per second. Therefore, the data may not have been fine enough to detect any differences.

Unlike previous research, ΔFT was found to have significant correlations with subjective workload measures (greater than 0.84, Table 4.52), indicating that ΔFT decreased while subjective workload decreased. Previous research investigated simple tasks, but in a complex task ΔFT may be indicative of mental workload, due to the types of mental activities that are occurring. These results were similar with overall adjustment session (0.99 with SWAT, Table 3.7) and low performers adjustment session (0.99 with MCH, Table 3.21) in chapter 3. An adjustment session required that participants performed at least 2 tasks, solving questions and adjusting their subjective response time at the same time and this study required complex tasks such as driving. Because ΔFT was not correlated with performance, it is difficult to conclude that ΔFT is related to mental workload. Further study is needed to more completely understand this finding.

Table 4.51

Correlation coefficients between performance, subjective workload measures, and thermography for novice drivers

Performance	MCH	SWAT	ΔNT
Collision	**0.89**	**0.93**	**-0.88**
Lost control	NS	NS	**-0.92**
Perception time	NS	NS	NS
Movement time	NS	NS	NS
Total brake time	NS	NS	NS
Parking duration (City)	**0.99**	**0.97**	**-0.90**
Parking failure (City)	**0.98**	**0.97**	**-0.89**
Lane deviation (Highway)	NS	**0.90**	NS
S-Curve (Rural)	NS	**0.82**	**-0.89**
Lane deviation (Rural)	NS	NS	NS

NS denotes no significant effect

Table 4.52

Correlation coefficients between subjective workload measures and thermography for novice drivers

	MCH	SWAT
ΔNT	-0.83	-0.87
ΔFT	0.88	0.84

Hypothesis 5: Facial temperature readings will be constant after learning has occurred.

The findings for this study do not provide support for this hypothesis. However, thermography trends were consistent with performance trends for novice and experienced drivers. For both driver groups, an increasing linear trend was found for nose temperature over the course of the testing session. Experienced drivers showed no significant difference between stages, indicating that the increase was minimal (Table 4.54). Novice drivers had thermal readings in stage 1 that were significantly lower than stages 5 and 6, indicating that most skills were learned before stages 5 and 6. It was difficult to determine a sufficient overall training time for novice drivers because skills were mastered at different rates due to different skill difficulty level (Table 4.53). Additional training time likely was needed for a number of performance measures.

Table 4.53

Sufficient training time for each performance

Performance	Training stage	Performance	Training stage
Perception time	More than 6 stages	Parking duration	1 and 2
Movement time	No need	Parking failure	1 and 2
Total brake time	More than 6 stages	Lane deviation (Highway)	1 and 2
Collision	1 and 2	Lane deviation (Rural)	No need
Lost control	1	S-Curve	No need

Table 4.54

Adapting time for each performance for experienced drivers

Performance	Adapting stage	Performance	Adapting stage
Perception time	No need	Parking duration	1
Movement time	No need	Parking failure	1
Total brake time	No need	Lane deviation (Highway)	No need
Collision	1	Lane deviation (Rural)	No need
Lost control	No need	S-Curve	No need

Hypothesis 6: There will be no gender differences in mental workload or performance.

For all experienced and novice drivers, no gender main effect was found in this study for any of the dependent variables, indicating that males and females had similar mental workload and performance levels (Table 4.1 and Table 4.2). Experience by gender interaction effects were found for ΔNT and parking duration. No pairwise differences in ΔNT were found in the post hoc analysis. However, parking duration showed that experienced male drivers performed better than male and female novice drivers, while experienced female drivers had similar performance level as experienced male and novice female drivers (Table 4.3). These findings indicated that males performed better once they mastered the parking skill, but performed worse than females before learning the skill.

For experienced drivers, ΔNT showed that females experienced higher mental workload than males, indicating that females stressed more than males when adapting to a new working environment (Table 4.6). Gender differences were also found for parking performance (parking duration and failure), indicating that females needed more time and had higher failure rate than males although females had an experience from real car (Table 4.10). A gender by stage interaction effect was found for MCH rating for highway scenarios, indicating that females experienced higher mental workload than males that was consistent with ΔNT (Table 4.15).

For novice drivers, gender difference was not found for any of the dependent variables, nor the interaction effects. This finding indicated that males and females had similar performance level and mental workload when they learned a new task.

For both experienced and novice drivers, females' ΔNT was lower than males' and subjective workload ratings showed that the females' average was higher than the males', indicating that females experienced a higher mental workload than males on average (Table 4.55).

Females appeared to drive more cautiously because they performed better than males in critical performance activities such as braking event, lost control, S-curve, and collision in general, indicating that females have lower accident rate than male (Sanders and McCormick, 1993).

Table 4.55

Mean for gender in each variables for experienced and novice drivers

Variable	Female	Male	Variable	Female	Male
ΔNT	-0.75	0.28	Collision	0.84	1.00
MCH	2.76	2.66	Parking duration	64.5	62.36
SWAT	131.4	118.3	Parking failure	0.84	0.78
Perception Time	0.50	0.51	Land deviation (highway)	1.20	1.19
Movement time	0.45	0.50	S-Curve	0.26	0.27
Total brake time	0.95	1.00	Land deviation (Rural)	0.69	0.71
Lost control	0.19	0.22	Blinker	0.28	0.30
Over-speed	3.93	4.86	Average speed	41.17	41.73

4.9 FUTURE WORK AND LIMITATIONS

This study showed that different tasks required different learning times. Also, imposed mental workload could be different between performances. To investigate learning progression in complex tasks, researchers need to assess each skill independently to tailor training to individual needs, using part-task training (Wickens et al, 1998; Wickens and Hollands, 2000; Wightman and Lintern, 1985). While this will require additional work for researchers or trainers, it will ensure sufficient training for skill mastery of critical skills. Previous research has shown that when training involved hands on activities with immediate feedback, performance and learning occur at an increased rate (Salas et al., 2006). Building upon skills can allow for sufficient training on skills sets without increasing training time on individual skills.

This study used experienced drivers aged 18 to 25 years, with an average of 20.5 years. This age range may have been too low to control impulsive behaviors while driving in the simulator. Future studies should include experienced drivers from a wider range of ages to allow for the incorporation of drivers with proven lower accident rates (Sanders and McCormick, 1993; Aizenberg and Mckenxie, 1997). In this manner, differences between age groups and experience levels can be assessed.

To improve studies in learning progress, individual differences between individuals should be more closely examined. Each individual needs to investigate whether specific individual learning styles (fast or slow learner) may influence learning and skill mastery (as found in chapter 3). A longer experiment (more days) needs to be conducted or learning performance numbers need to be reduced to investigate the relationship between overall performance and thermography in learning progression.

Thermal readings for this data were collected every 10 seconds. This led to the loss of data, particularly for events that had durations less than 10 seconds (such as perception time, movement time, and total brake time). Additional data collection may provide a more complete and accurate picture of thermal readings during complex task performance.

Future studies should consider the fatigue and its effects on mental workload and performance. This study observed that participants were not concerned about other factors (fatigue, stress, or boredom) during reporting of subjective workload ratings.

4.10 CONCLUSIONS

This study provides preliminary support for the use of thermography to quantify mental workload during complex task performance and training. Thermography was strongly and significantly correlated with subjective mental workload measures and some performance measures. Further research is needed to understand mental workload and thermographic fluctuations during performing complex tasks and during training. The role of multi-tasking on thermal images needs to be understood more fully to assist in quantifying sufficient training times for more complex tasks.

4.11 REFERENCES

Andre, A.D., & Wickens, C.D. (1995). When users want what's not best for them. *Ergonomics in Design:The Quarterly of Human Factors Applications 3(4)*, Oct. 10-14(5).

American Society for Training and Development (ASTD) (2003). *ASTD 2003 State of the Industry Report Executive Summary*, Retrieved June 3, 2004, From http://www.astd.org/NR/rdonlyres/6EBE2E821D2948A78A3A%20357649BB6DB6/0/SOIR_2003_Executive_Summary.pdf.

Alicandri, E. (1994). The highway driving simulator (HYSIM): The next best thing to being on the road. *Public Roads*, 57, 19-23.

Allen, R.W., Stein, C.A. Aponso, B.L., Rosenthal, T.J., & Hogue, J.R. (1990). A low cost, part task driving simulator based on microcomputer technology. *Proceedings of the Paper presented at the 69th Annual Meeting of the Transportation Research Board*, Washington, DC.

Allen, R.W., Park, G., Cook, M., Rosenthal, T.J., Riorentino, D., & Viirre, E. (2003). Novice driver training results and experience with a PC based simulator. *Proceedings of the second International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. Park City, UT, 165-170.

Aizenberg, R. & McKenzie, D.M. (1997). Teen and senior drivers, *California Department of Motor Vehicles Report No. CAL-DWV_RSS-97-168*, December 1997.

Carsten, O.M.J., Groeger, J.A., Blana, E., & Jamson, A.H. (1997). *Driver performance in the EPSRC driving simulator: A validation study* (Report No. GR/K56162). United Kingdom: University of Leeds.

Catrambone, R., & Carroll, J.M. (1987). Learning a word processing system with training wheels and guided exploration. *Proceedings of CHI & GI human-factors in computing systems and graphics conference.* New York: Association for Computing Machinery, 169-174.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.).* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Collet, C., Petit, C., Champely, S., & Dittmar, A. (2003). Assessing workload through physiological measurements in bus drivers using an automated system during docking. *Human Factors*, 45, 539-548.

Colquitt, J.A., LePine, J.A., & Noe, R.A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research.*Journal of Applied Psychology*, 85, 678-707.

Desmond, P.A.,& Matthews, G. (1997). Implication of task-induced fatigue effects for in-vehicle countermeasures to driver fatigue. *Accident Analysis and Prevention*, 29, 515-523.

Dorn, L., & Barker, D. (2005). The effects of driver training on simulated driving performance. *Accident Analysis and Prevention*, 37, 63-69.

Ehlert, P.A.M., & Rothkrantz, L.J.M. (2001). Microscopic traffic simulation with reactive driving agents. *Proceedings of the2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.01TH8585)*, Piscataway, NJ: IEEE: 860-865.

Ellingrod, V.L., Perry, P.J., Yates, R.W., Maclnodoe, J.H., Watson, G., Arnalt, S., & Holman, T.L. (1997). The effects of anabolic steroids on driving performance as assessed by the Iowa Driver simulator, *American Journal of Drug Alcohol Abuse,* 23, 623-637.

Federal Aviation Administration (FAA). (2005). *FAA Human Factors Awareness Course*. Retrieved December 7, 2005. From http://wfnetwork.bc.edu/glossary_entry.php?term=Workload,%20Definition(s)%20of&area=All

Fraser, D.A., Hawken, R.E., & Warnes, A.M. (1994). Effects of extra signals on drivers' distance keeping: a simulation study. *IEEE Transaction on VehicularTechnology*, 43, 1118-1124.

Fisher, D. L., Laure, N.E., Glaser, R., Connerney, K., Pollatsek, A., Duffy,S.A., et al. (2002). Use of a fixed-based driving simulator to evaluate the effects of experience and PC-based risk awareness training on drivers' decisions, *Human Factors*, 44, 287-302.

Fisk, A.D., Ackerman, P.L., & Schneider, W. (1987). Auomatic and controlled processing theory and its applications to human factors problems. In P.A. Hancock (Ed.), *Human Factors Psychology* (pp. 159-197). Amsterdam: Elsevier.

Galinsky, T.L., Rosa, R.R., Warm, J.S., & Dember, W.N. (1993). Psychophysical determinants of stress in sustained attention. *Human Factors*, 35, 603-614.

Galvin, T. (2002). 2002 Industry Report, *Training*, 39, October, 24-73.

Genno, H.[a], Saijo, A., Yoshida, H., Suzuki, R., & Osumi, M. (1997). Non-contact method for measuring facial skin temperature. *International Journal of Industrial Ergonomics,* 19*,* 147-159.

Genno, H.[b], Ishikawa, K., Kanbara, O., Kikumoto, M., Fujiwara, Y., Suzuki, R., et al. (1997). Using facial skin temperature to objectively evaluate sensations. *International Journal of Industrial Ergonomics,* 19, 161-171.

Godley, S. T. (2002). Driving simulator validation for speed research. *Accident Analysis Prevention*, 34, 589-600.

Green, J., & Shellenberger, R. (1991). *The dynamics of health and wellness: A Biopsychosocial Approach*. FortWorth: Holt, Rinehart and Winston, Inc.

Green, M. (2000). How long does it take to stop? Methodological analysis of driver perception-brake times. *Transportation Human Factors*, 2, 195-216.

Gregerson, N.P. (1996). Young car drivers: why are they over represented in car accidents? how can driver training improve their situation? (*VTI Rapport 409A*). Linkoping; Swedish National Road and Transport Institute.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: Elsevier Science Publishers B. V.

Harrell, C. R., Ghosh, B. K., & Bowden, R. O. (2004).*Simulation Using ProModel (2nd ed.)*. New York: McGraw-Hill.

Hunter, J.E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance.*Journal of Vocational Behavior*, 29, 340-362.

Insurance Institute for Highway safety (2003). Fatality facts, older people. http://www.hightwaysafety.org/safety_facts/fatality_facts/olderpeople.htm(Accessed April 10, 2005).

Jacobs, J.W., & Dempsey, J.V. (1993). Simulation and gaming: Fidelity, feedback, and motivation, In J.V. Dempsey & G.C. Sales (Eds.), *Interactive Instruction and Feedback* (pp.197-229). Englewood Cliffs, NJ: Educational Technology.

Johansson, G., & Rumar, K. (1971). Drivers' brake reaction time. *Human Factors*, 13, 23-27.

Johnson, D.F., & Haygood, R.C. (1984). The use of secondary tasks in adaptive training. *Human Factors*, 26, 105-108.

Joyce, M. (1999). The role of ergonomics training in industry, In W.Karwowski & W.S. Marras (Eds.), *The Occupational Ergonomics Handbook* (pp.1631-1640). New York: CRC Press.

Kahneman, D. and Tversky, A. (1973) On the psychology of prediction, *Psychology Review*, 80, 237-251

Kang, J., McGinley, J., McFadyen, G., &Babski-Reeves, K. (2006). Determining learning level and effective training times using thermograhy, *Proceedings of 25th Army Science Conference*, Orlando, FL.

Kang, J., & Babski-Reeves, K. (2008). Detecting mental workload fluctuation during learning a novel task by thermography. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual meeting* , September 22-26, 2008, New York, NY.

Lee, H.C., Cameron, D., & Lee, A.H. (2003). Assessing the driving performance of older adult drivers: on-road versus simulated driving. *Accident Analysis and Prevention*, 35, 797-803.

Liebermann, D. G., Ben-David, G., Schweitzer, N., Apter, Y., & Parush, A. (1995). A field study responses during driving.; I. Triggering and modulation. *Ergonomics*, 38, 1894-1902.

Mayhew, D.R., & Simpson, H.M. (1995). The role of driving experience: Implications for training and licensing of new drivers. *Insurance Bureau of Canada*: Toronto, Canada.

McDonald, N. (1984). *Fatigue, safety, and the truck driver*. Taylor & Francis, London.

McKnight, J.A., & McKnight, S.A. (2003). Young novice Drivers: Careless or clueless. *Accident Analysis and Prevention*, 35, 921-925.

Nilsson, L.(1993). Behavioural research in an advanced driving simulator-experiences of the VTI system.*Proceedings of the Human Factors and Ergonomics society 37th Anrual Meeting,* 37, 612-616.

Nilsson, T, Nelson, T.M., and Carlson, D. (1997). Development of fatigue symptoms during simulated driving. *Accident Analysis, and Prevention*, 29, 479-488.

Pradhan, A.K., Fisher, D.L., &Rollatsek, A. (2005). The effects of PC-based training on novice drivers' risk awareness in a driving simulator, *Proceedings of the third International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Rockport, ME. USA,81-87.

Randel, J.M., Main, R.E., Seymour, G.E., & Moris, B.A. (1992). Relation of study factors to performance in Navy technical schools. *Military Psychology*, 4, 75-86.

Ranney, T.A. (1994). Models of driving behavior: A review of their evaluation. *Accident Analysis and Prevention*, 26, 733-750.

Rasmussen, J. (1983). Skills, rules, knowledge: Signals, signs, and symbols and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13, 257-267.

Rasmussen, J (1986). *Information processing and Human-machine interaction: An approach to cognitive engineering*. New York: Elsevier.

Rasmussen, J (1993). Deciding and doing: Decision making in natural contexts, In G.Klein, J, Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 158-171). Norwood, NJ: Ablex.

Ree, M.J., & Earles, J.A. (1991). Predicting training success: Not much more than G. *Personality Psychology,* 44, 321-332.

Ree, M.J., Carretta, T.R., & Teachout, M.S. (1995). Role of ability and prior job knowledge in complex training performance. *Journal of applied psychology*, 80, 721-730.

Richter, R.L., & Hyman, W.A. (1974). Research note: Driver's brake reaction times with adaptive controls. *Human Factors*, 16, 87-88.

Roenker, D.L. & Cissell, G.M. (2003). Speed-of-processing and driving simulator training result in improved driving performance. *Human Factors*, 45, 218-233.

Rogers, W.A., Rousseau, G.K., & Fisk, A.D. (1999). Applications of attention research. In F.T. Durso (Ed.), *Handbook of Applied Cognition* (pp. 33-56). New York: John Wiley & Sons.

Ross, L., Lepper, M. R. and Hubbard, M. (1975) Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm, *Journal of Personality and Social Psychology,* 32, 880-892

Roscoe, A.H. (1992). Assessing pilot workload. Why measure heart rate, HRV, and respiration?. *Biological Psychology*, 34, 259-287.

Sales, E., & Cannon-Bowers, J.A. (2001). The Science of Training: A Decade of Progress. *Annual Review of Psychology*, 52, 471-499.

Salas, E., Wilson, K.A., Priest, H.A., & Guthrie, J.W. (2006). Design, Delivery, and evaluation of training systems, In G. Salvendy (Ed.). *Handbook of Human Factors and Ergonomics (3rd ed.)* (pp. 472-512). Hoboken, NJ: John Wiley& Sons, Inc.

Sanders, M.S., & McCormick E., J. (1993). *Human factors in engineering and design (7th ed.)*. New York, NJ: McGRAW-Hill, Inc..

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. *Human Factors*, 27, 285-300.

Skipper, J.M., & Wierwille, W.W. (1986). Drowsy driver detection using discriminant analysis. *Human Factors*, 28, 527-540

Storie, V.J. (1984). Involvement of goods vehicles and public service vehicles in motorway accidents. *UK Department of Transport/Transport and Road Research Laboratory*, Report No. 1113.

Stoll, A.M. (1964). Techniques and uses of skin temperature measurements. *Annals of The New York Academy of Sciences: Thermography and Its Clinical Applications*, 121(1), 49-56.

Summala, H., Nieminen, T., & Punto, M. (1996). Maintaining lane position with peripheral vision during in-vehicle tasks. *Human Factors*, 38, 442-451.

Swezey, R.W., & Pearlsten, R.B. (2001). Selection, training, and development of personnel. In G. Salvendy (Ed.). *Handbook of Industrial Engineering: Technology and Operations Management (3rd ed.)* (pp 920-947). New York, NY: John Wiley & Sons, Inc.

Te Velde, A.F., der Kamp, J.V., Barela, J.A., & Savelsbergh, G.J.P. (2005). Visual timing and adaptive behavior in a road-crossing simulation study. *Accident Analysis and prevention,*37, 399-406.

Trujillo, A.C. (1998). Pilot mental workload with predictive system status information. *Proceedings of the 4th Symposium on Human Interaction with Complex Systems*, Fairborn, OH,73-80.

Van der Winsum, W. (1996). Speed choice and steering behaviour in curve driving. *Human Factor*, 38, 343-351.

Van der Winsum, W., & Brouwer, W. (1997). Time headway in car following and operationsl performance during unexpected braking. *Perceptual and Motor Skills*,84, 1247-1257.

Veltman, H.J.A., & Vos, W.K. (2005). A methodology for non-intrusive mental workload measurement in simulated and on-road driving. *Proceedings of the11th International conference on Human-computer interaction*, Las Vegas, NV.

Vicente, K.J. (1992). Memory recall in a process control system: A measure of expertise and display effectiveness.*Memory & Cognition*, 20, 356-373.

Wickens, C.D., Gordon, S.E., & Liu, Y. (1998). *An Introduction to Human Factors Engineering*. New York: Addison Wesley Longman.

Wickens, C.D., & Hollands, J.G. (2000). *Engineering Psychology and Human Performance (3rd ed.)*. Upper Saddle River, NJ: Prentice_Hall.

Wightman, D.C., & Lintern, G. (1985). Part-task training for tracking and manual control. *Human Factors*, 27, 267-283.

Winsum, W.V., & Brouwer, W. (1997). Time headway in car following and operational performance during unexpected braking. *Perceptual and Motor Skills*, 84, 1247-1257.

Ye, N., & Salvendy, G. (1994). Quantitative and qualitative differences between experts and novices in chunking computer software knowledge. *International Journal of Human-Computer Interaction*, 6, 105-118.

CHAPTER 5

CONCLUSION

Thermography has been illustrate to be a valid, objective, and non-invasive mental workload measure, in particular for simple, novel tasks, and in identifying "optimal" workload levels. Thermography has the capability to demonstrate mental workload level profiles during segments of a tasks, a limitation of many current subjective workload measures and performance measures. Subjective and performance measures cannot demonstrate workload level fluctuation during performing a task continuously, due to their design; however, thermography is able to capture low level changes in workload in real time.

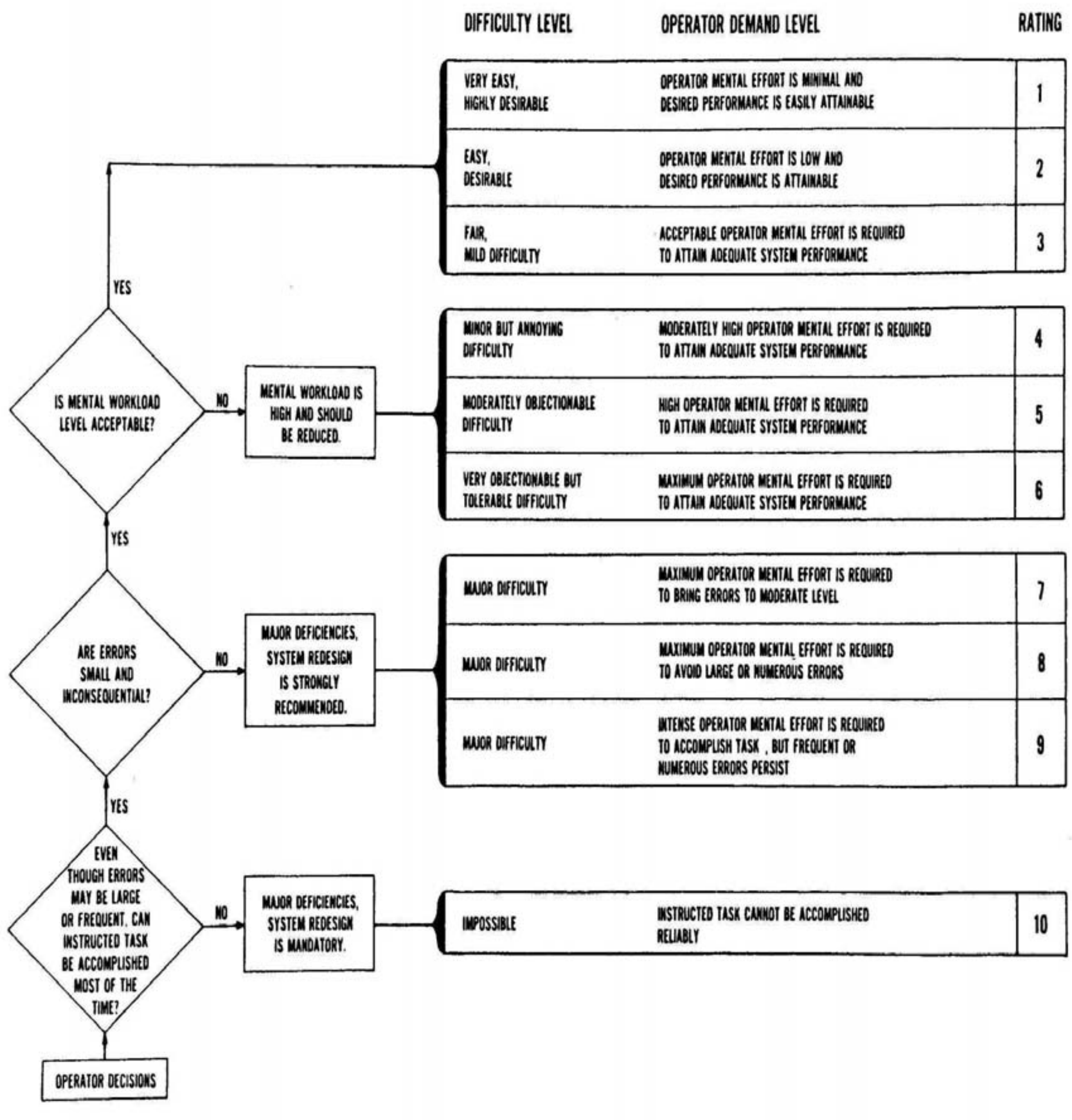
Further, thermography illustrates that nose temperature increases during skill acquisition, which is useful in assessing training times. Also, changes in forehead temperature may be associated with executive mental processes, and may be useful in assessing the workload assessed by multi-tasking. Additional studies are needed to fully understand the utility of using thermography to design training systems and the effects of multi-tasking on thermal readings.

Thermography provides a more practical mental workload assessment tool for a couple of reasons. First, thermography does not require the interruption of the task or the attachment of equipment for physiological response measurement. Second, changes in the thermal readings can be assessed in real time. Third, thermography is not prone to individual bias and expectations (as is the case in subjective assessments). Fourth, thermography is not prone to the same limitations as performance measures of mental workload, though thermography is correlated with some performance metrics.

In general, the results from this research support the continued use of thermography as a mental workload assessment technique. Additional research on various types of tasks, protocols for identifying optimal workload levels, and changes in forehead temperature as a result of mental processing tasks is needed. Further, the influence of common environmental factors (such as air conditioning, extreme temperatures, etc.) and sophisticated image processing techniques are needed to improve the utility of thermographic assessments in real time.

APPENDIX A

MODIFIED COOPER-HARPER SCALE



The Modified Cooper-Harper scale (Wierwille, W. W. and Casalli, J. G., 1983).

APPENDIX B

SUBJECTIVE WORKLOAD ASSESSMENT TECHNIQUE (SWAT)

Time Load, Mental Effort Load, and Stress Load Ratings

Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.	Occasionally have spare time. Interruptions or overlap among activities occur frequently.	Almost never have spare time Interruptions or overlap among activities are very frequent, or occur all the time.
Very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.	Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability, or unfamiliarity. Considerable attention required.	Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.
Little confusion, risk, frustration, or anxiety exists and can be easily accommodated.	Moderate stress due to confusion, frustration, or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.	High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self-control required.

APPENDIX C

DEMOGRAPHIC AND HEALTH SCREENING FORM

Demographic and Health Screening Form

Participant Data

Participant Number ____________________________ **Date** ____________________

Starting Time ____________________________ **Age** ____________________

Gender: Male Female

Do you have a valid U.S. Driver License ____________________________

Year of driving experience ____________________________

Race

Caucasian_______________ **Asian**________________

Hispanic_______________ **Pacific Islander**___________

African-American____________ **Other**_________________

Health Screen

1) Do you have any of the following conditions? Check all that apply

____Epilepsy

____Pregnancy

____Heart Problem

2) Are you feeling well today? ________

Vision

Do you have 20/20 or corrected to 20/20 vision or better? Yes____ No____

____Contacts

____Glasses

Are you color blind?   Yes____ No____

Primary foot to operate the brake

Which foot do you use on the brake?_________

APPENDIX D

PRE-EXPOSURE SIMULATOR SICKNESS FORM

Simulator Sickness Form

Pre-exposure Simulator Sickness Questionnaire

Pre-exposure instructions: please fill in this questionnaire. Circle below if any of the symptoms apply to you now. You will be asked to fill this again after the experiment

No symptom Severe

General discomfort	0	1	2	3
Fatigue	0	1	2	3
Headache	0	1	2	3
Eyestrain	0	1	2	3
Difficulty focusing	0	1	2	3
Sweating	0	1	2	3
Nausea	0	1	2	3
Difficulty concentrating	0	1	2	3
Fullness of head	0	1	2	3
Blurred vision	0	1	2	3
Dizzy (eyes open)	0	1	2	3
Dizzy (eyes closed)	0	1	2	3
Vertigo	0	1	2	3
Burping	0	1	2	3

APPENDIX E

POST-EXPOSURE SIMULATOR SICKNESS FORM

Simulator Sickness Form

Post-exposure Simulator Sickness Questionnaire

Post-exposure instructions: please fill in this questionnaire. Circle below if any of the symptoms apply to you now.

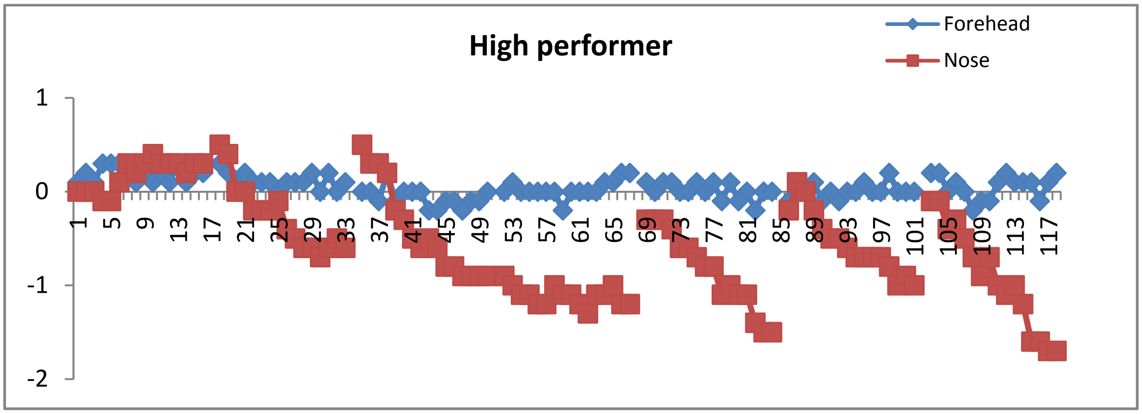
No symptom Severe

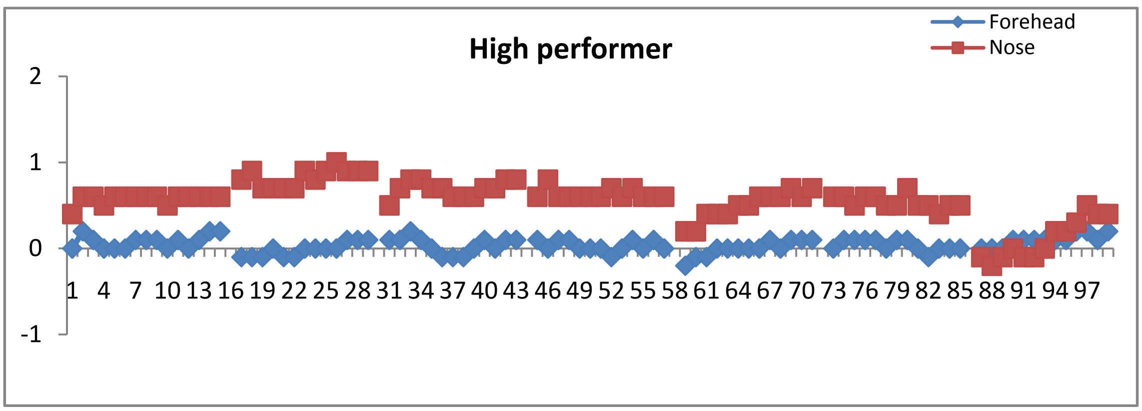
General discomfort	0	1	2	3
Fatigue	0	1	2	3
Headache	0	1	2	3
Eyestrain	0	1	2	3
Difficulty focusing	0	1	2	3
Sweating	0	1	2	3
Nausea	0	1	2	3
Difficulty concentrating	0	1	2	3
Fullness of head	0	1	2	3
Blurred vision	0	1	2	3
Dizzy (eyes open)	0	1	2	3
Dizzy (eyes closed)	0	1	2	3
Vertigo	0	1	2	3
Burping	0	1	2	3

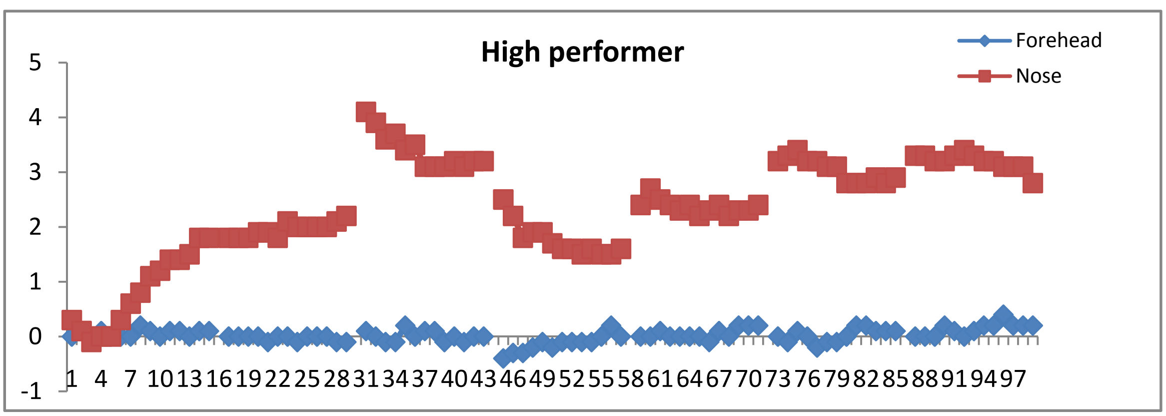
APPENDIX F

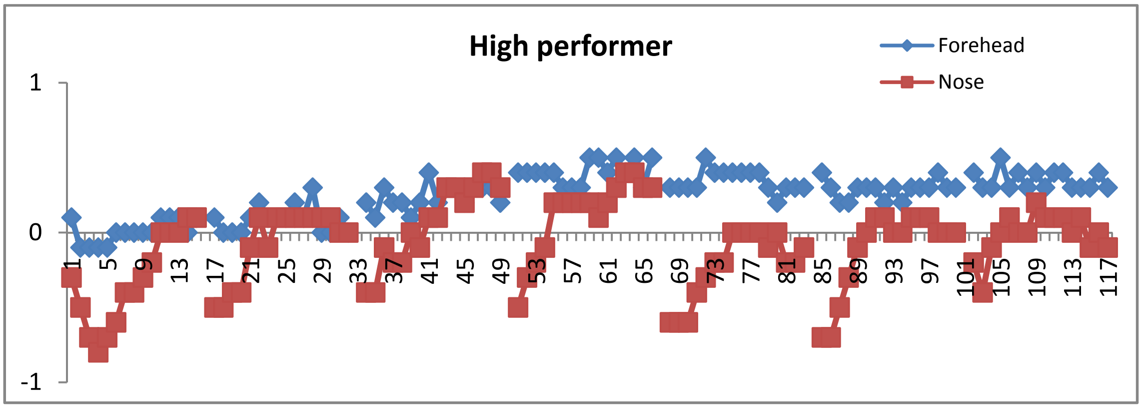
HIGH PERFORMERS' THERMAL READINGS

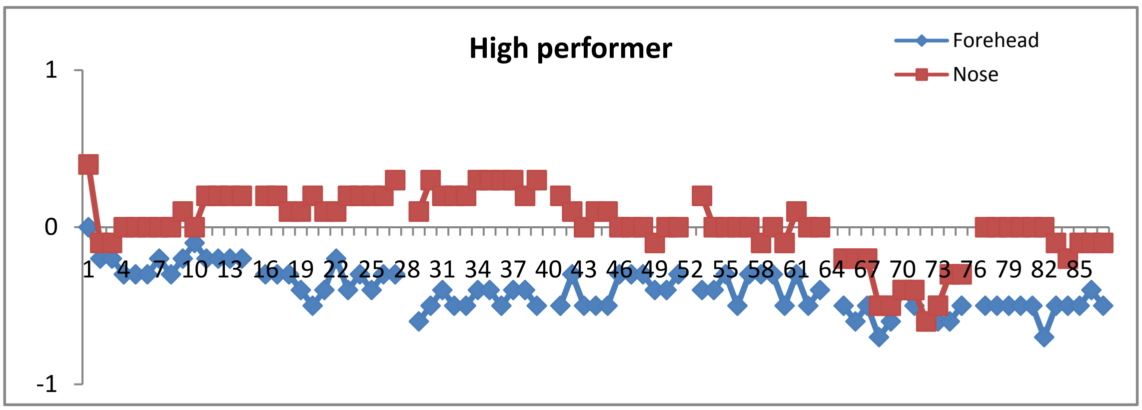
High performers' thermal readings

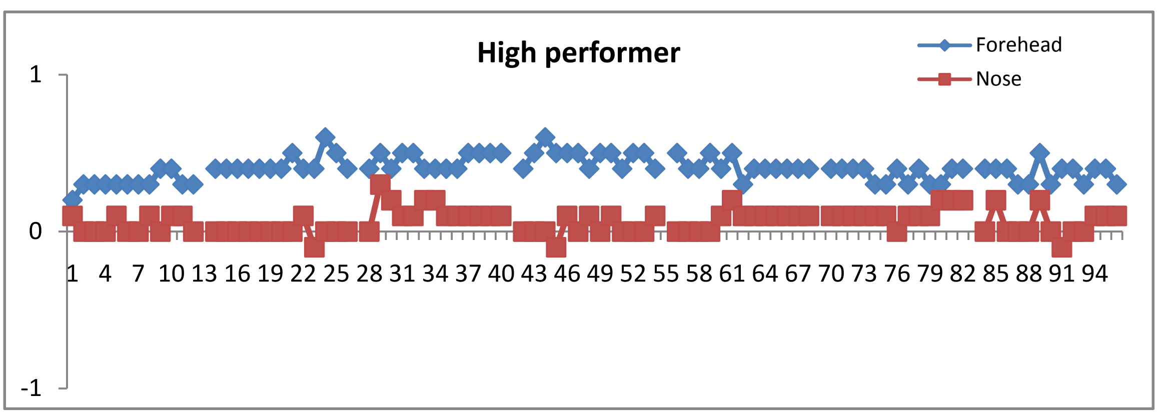


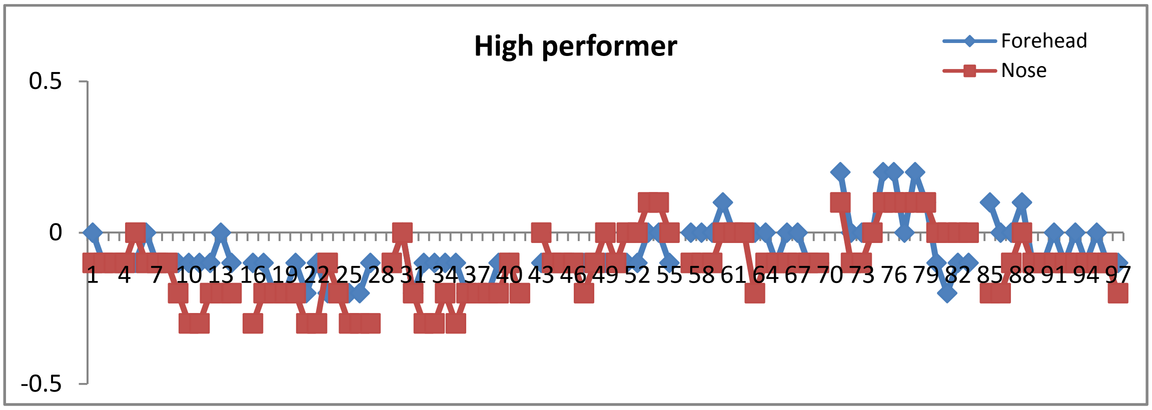


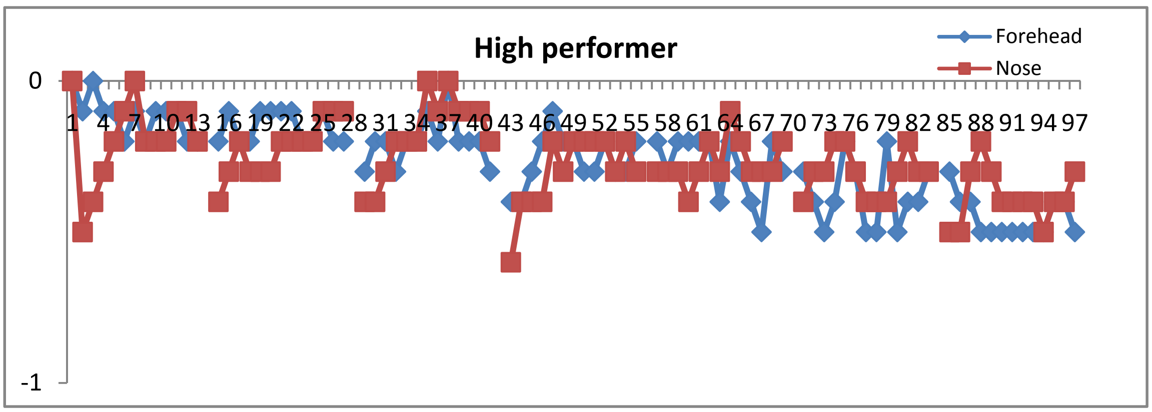


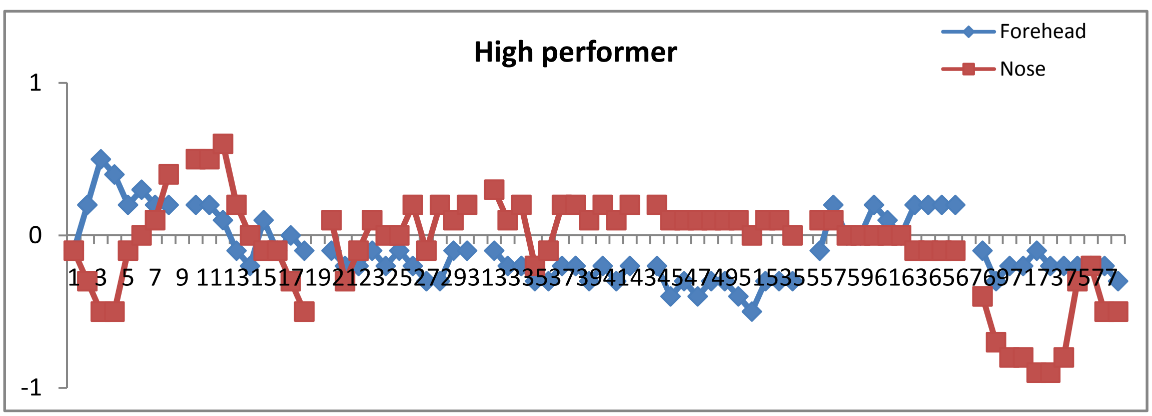

High performer
Forehead
Nose
1
0
-1
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89 93 97 101 105 109 113 117

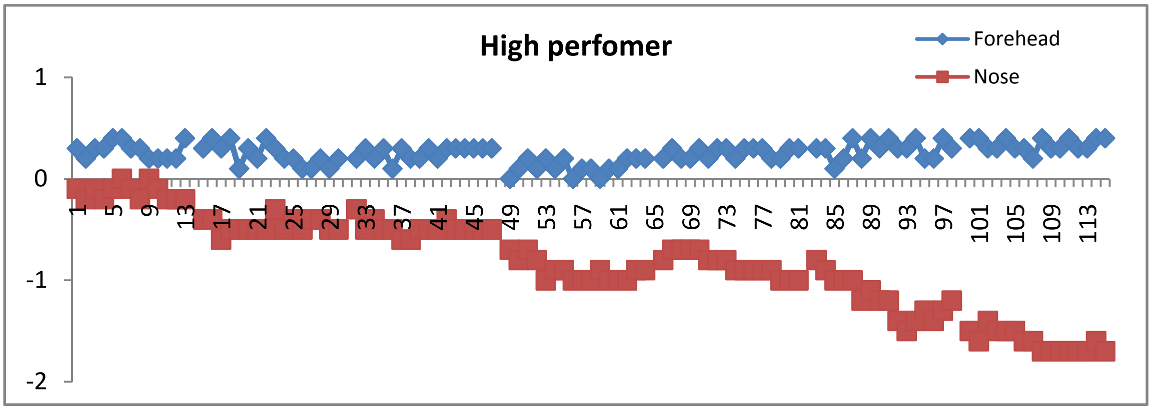

High performer
Forehead
Nose
1
0
-1
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85

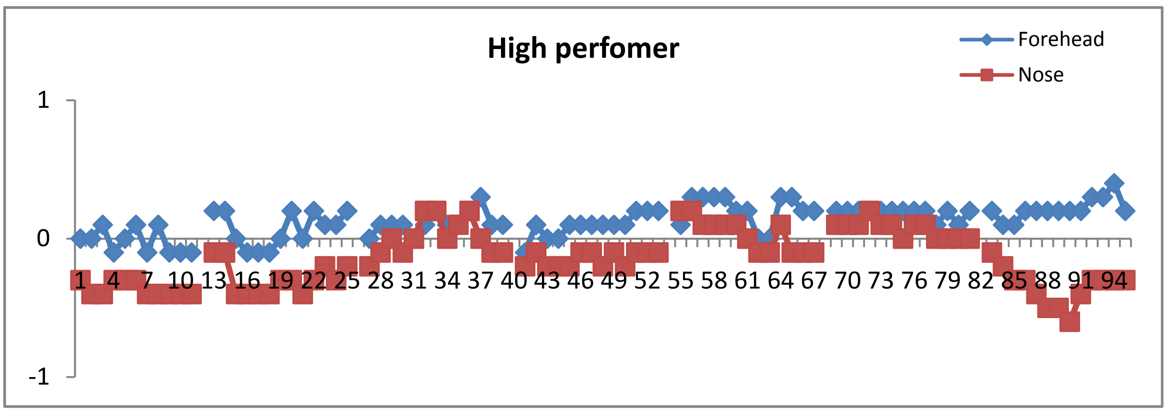

High performer
Forehead
Nose
1
0
-1
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85 88 91 94

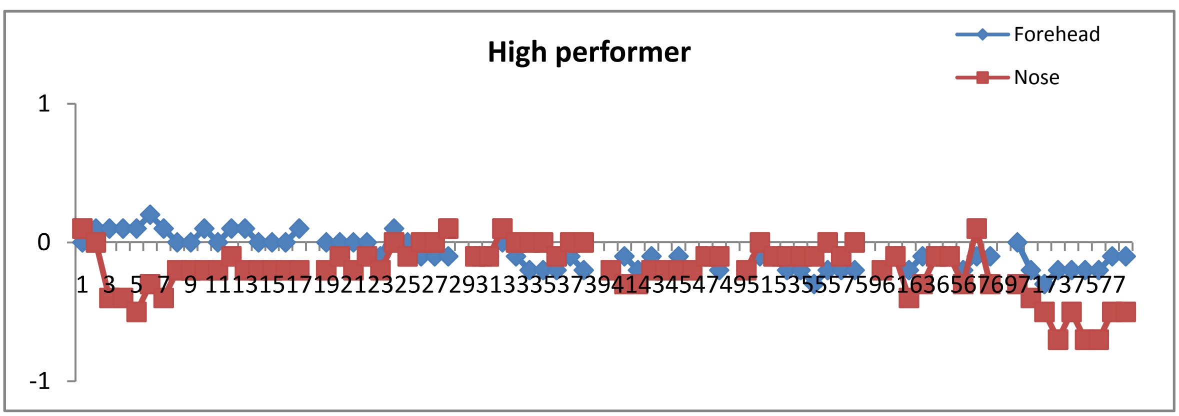

High performer
Forehead
Nose
0.5
0
-0.5

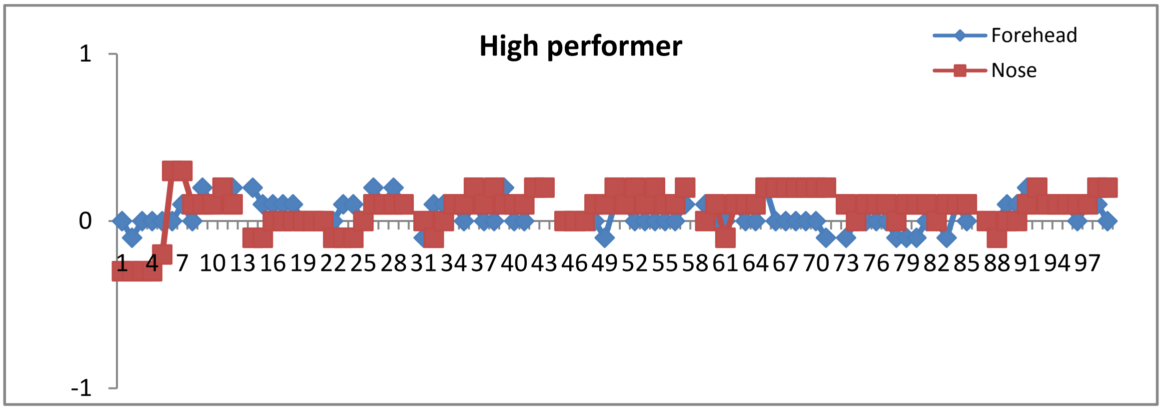

High performer
Forehead
Nose
0
-1

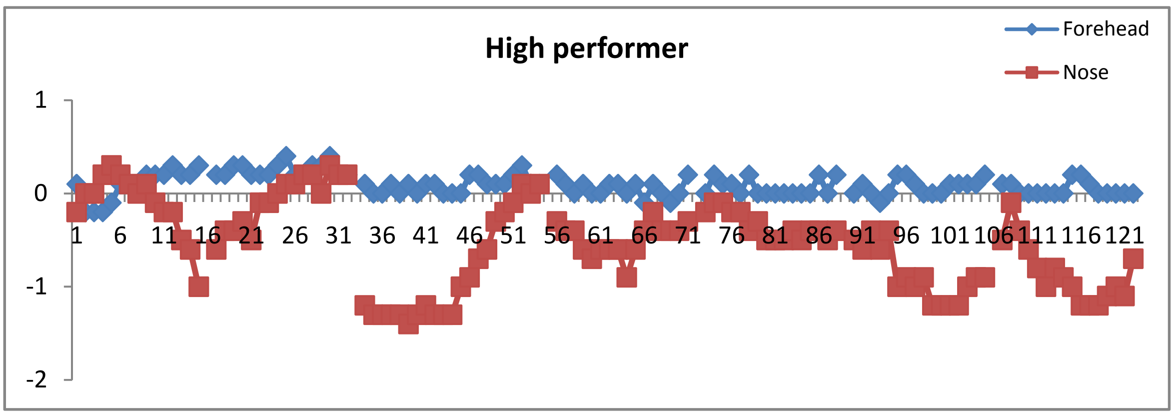

High performer
Forehead
Nose
1
0
-1

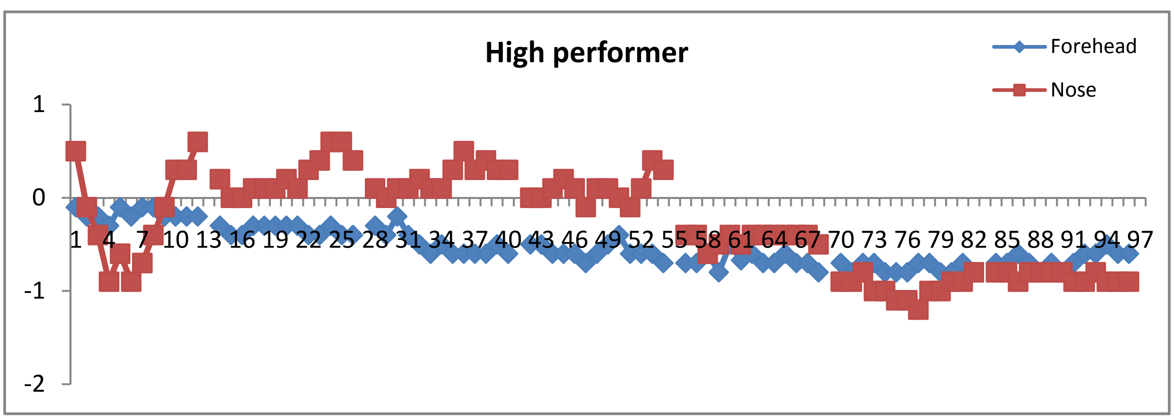

High perfomer
Forehead
Nose
1
0
-1
-2

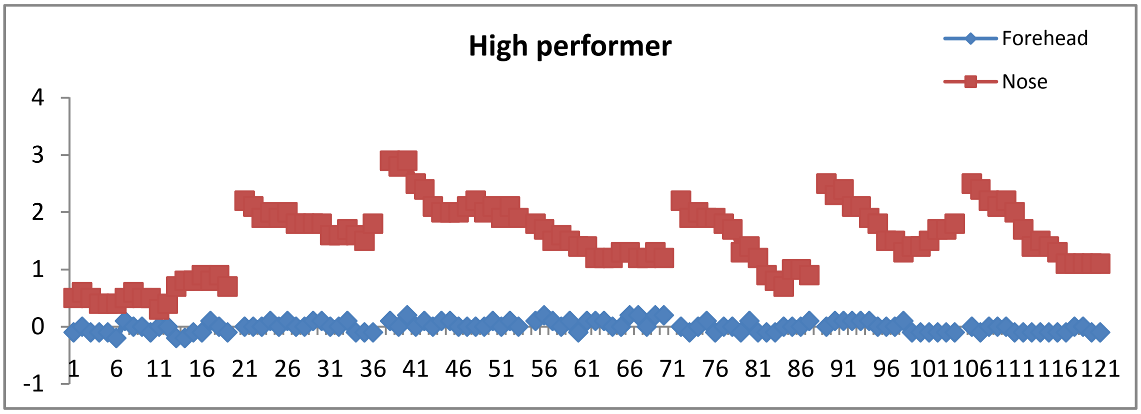

High perfomer
Forehead
Nose
1
0
-1

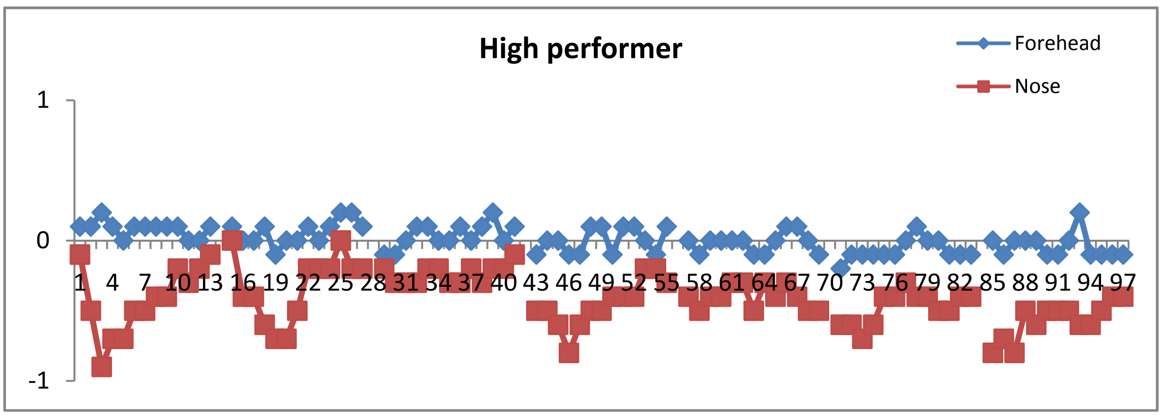

High performer
Forehead
Nose
1
0
-1

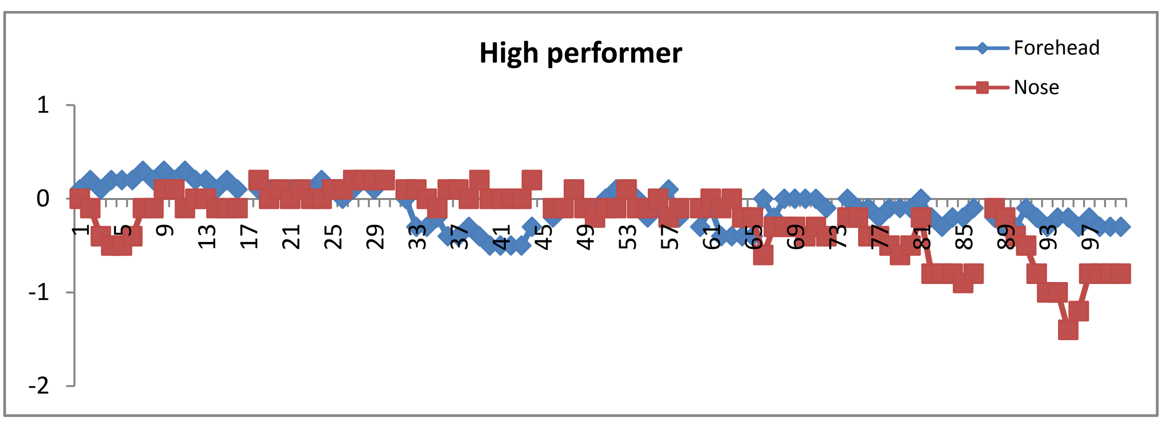

High performer
Forehead
Nose
1
0
-1
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85 88 91 94 97

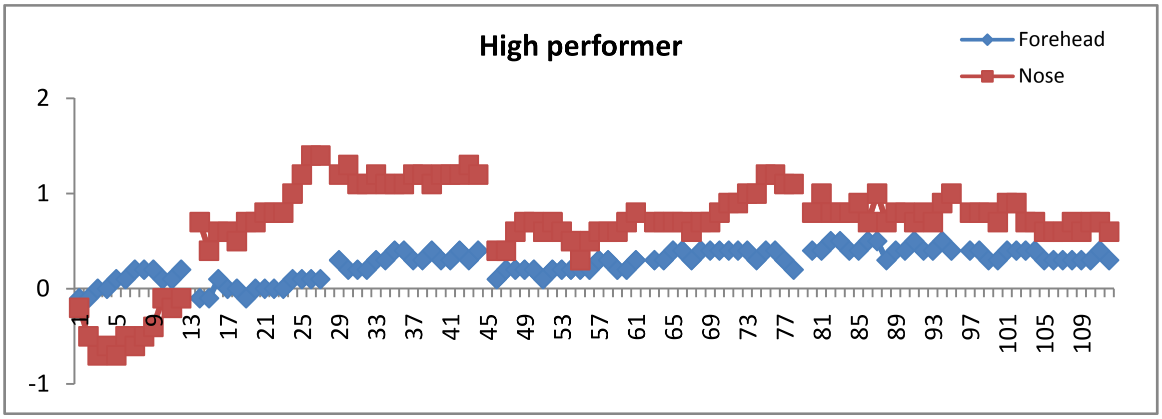

High performer
Forehead
Nose
1
0
-1
-2
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 111 116 121

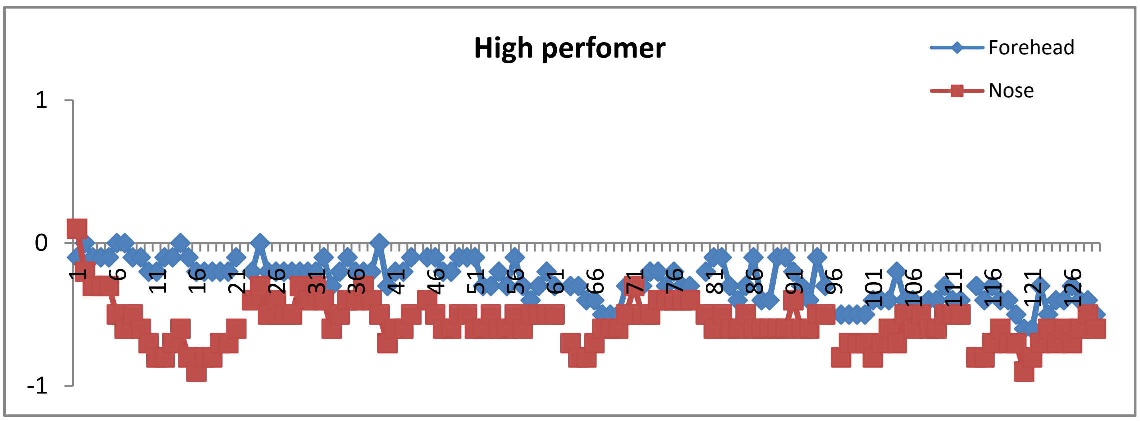

High performer
Forehead
Nose
1
0
-1
-2
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85 88 91 94 97


High performer
Forehead
Nose
4
3
2
1
0
-1
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 111 116 121


High performer
Forehead
Nose
1
0
-1
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85 88 91 94 97


High performer
Forehead
Nose
1
0
-1
-2
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89 93 97

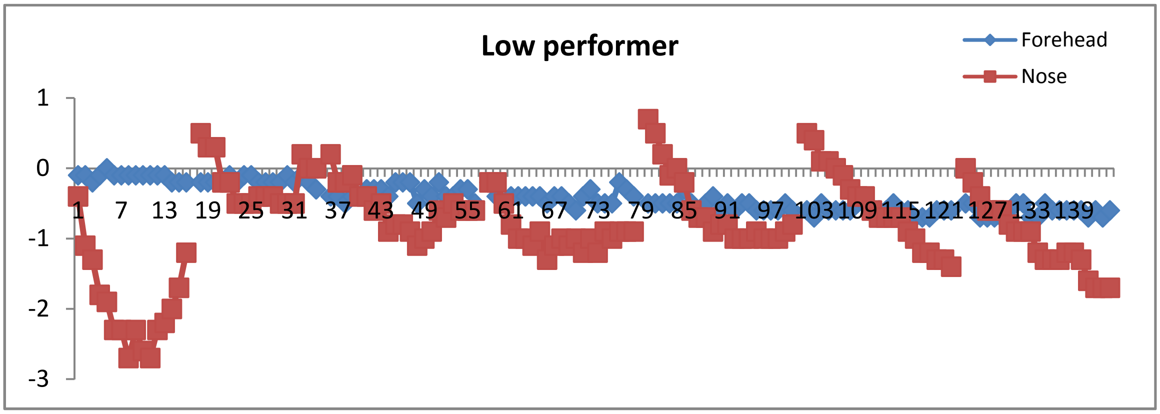

High performer
Forehead
Nose
2
1
0
-1
1
5
9
13
17
21
25
29
33
37
41
45
49
53
57
61
65
69
73
77
81
85
89
93
97
101
105
109

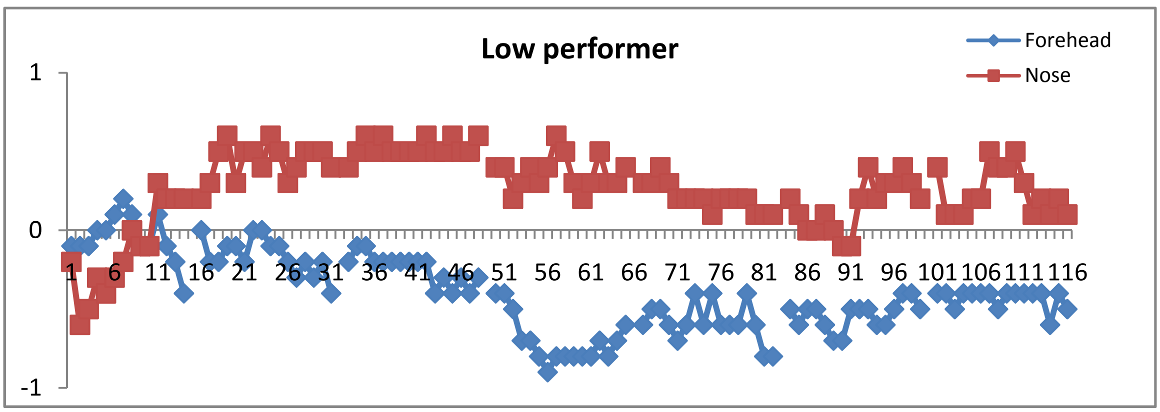

High perfomer
Forehead
Nose
1
0
-1
1
6
11
16
21
26
31
36
41
46
51
56
61
66
71
76
81
86
91
96
101
106
111
116
121
126

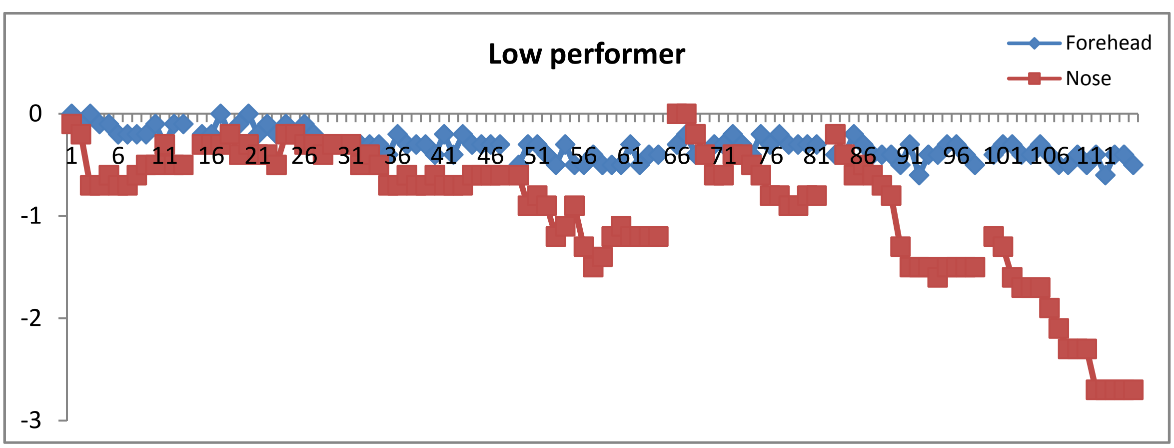
APPENDIX G

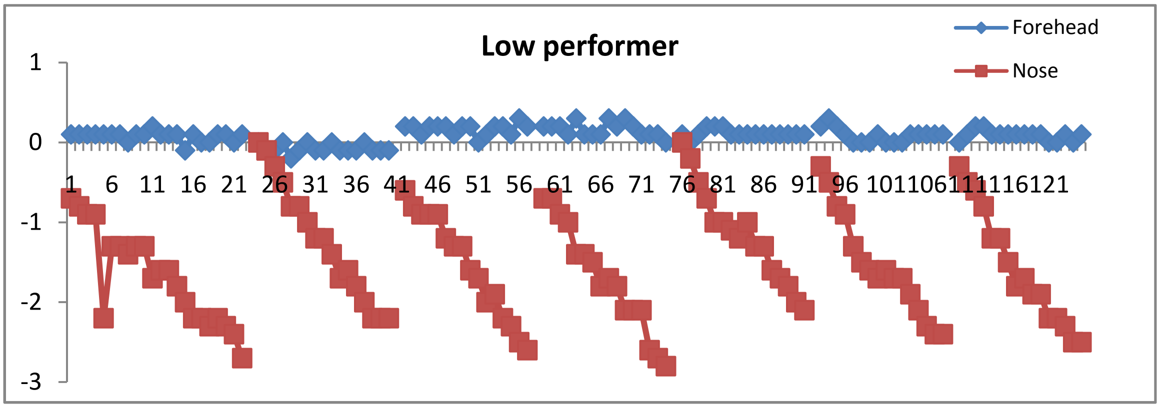
LOW PERFORMERS' THERMAL READINGS

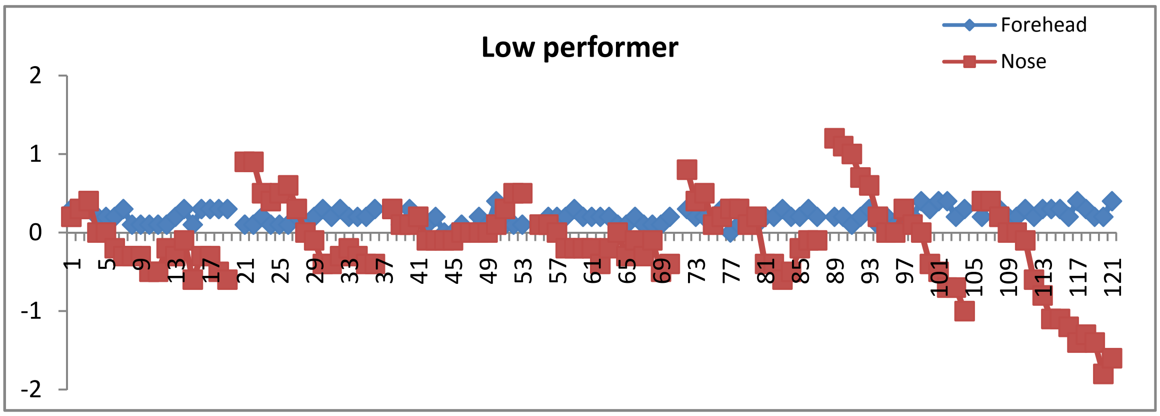
Low performers' thermal readings

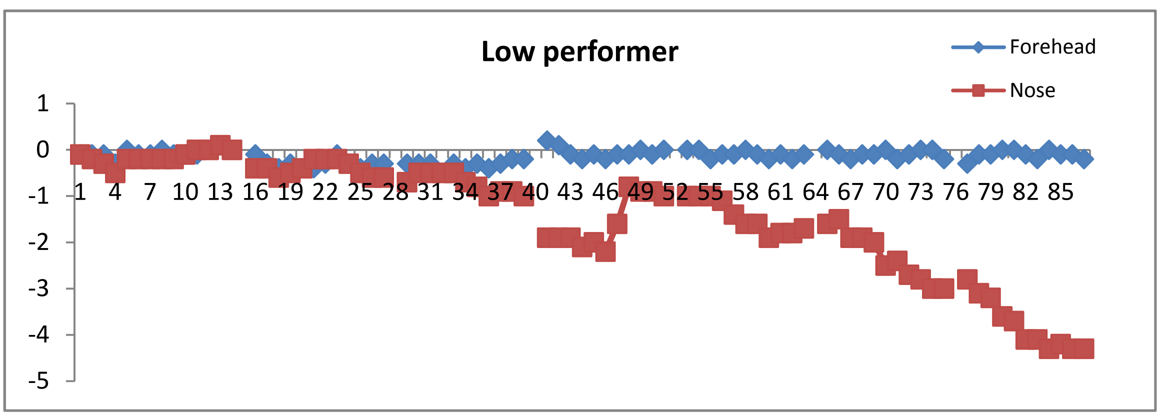


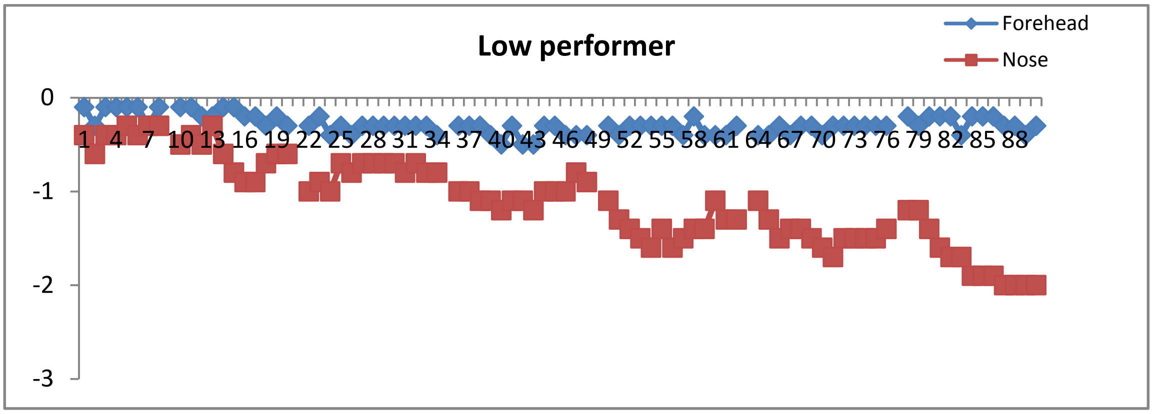


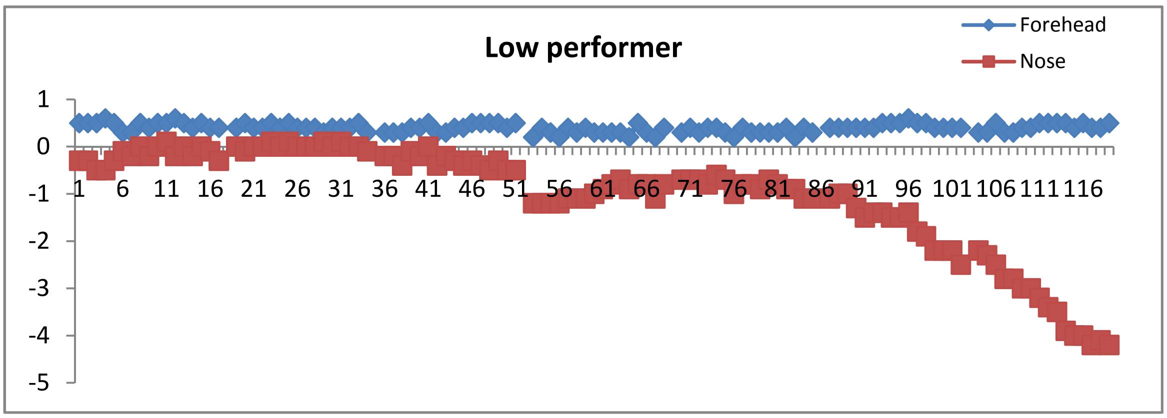



Low performer
Forehead
Nose
1
0
-1
-2
-3
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101106111116121


Low performer
Forehead
Nose
2
1
0
-1
-2


Low performer
Forehead
Nose
1
0
-1
-2
-3
-4
-5
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85


Low performer
Forehead
Nose
0
-1
-2
-3
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85 88


Low performer
Forehead
Nose
1
0
-1
-2
-3
-4
-5
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 111 116